

(12) INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(19) World Intellectual Property Organization
International Bureau



(43) International Publication Date
30 May 2002 (30.05.2002)

PCT

(10) International Publication Number
WO 02/42439 A2

- (51) International Patent Classification⁷: C12N 9/00
- (21) International Application Number: PCT/US01/49814
- (22) International Filing Date: 25 October 2001 (25.10.2001)
- (25) Filing Language: English
- (26) Publication Language: English
- (30) Priority Data:
60/243,916 27 October 2000 (27.10.2000) US
- (71) Applicant: GENETICS INSTITUTE, LLC [US/US]; 87 CambridgePark Drive, Cambridge, MA 02140 (US).
- (72) Inventors: RACIE, Lisa, Anne; 124 School Street, Acton, MA 01720 (US). TWINE, Natalie, Constance; 379 Shirley Hill Road, Goffstown, NH 03045 (US). AGOSTINO, Michael, John; 26 Wolcott Avenue, Andover, MA 01810 (US). WOLFMAN, Neil, Michael; 5 Phillips Lane, Dover, MA 02030 (US). MORRIS, Elisabeth, Ann; 15 Apple Street, Sherborn, MA 01770 (US).
- (74) Agents: GYURE, Barbara, A. et al.; c/o Kay E. Brady, American Home Products Corporation, Patent Law Department, Five Giralda Farms, Madison, NJ 07940 (US).
- (81) Designated States (*national*): AE, AG, AL, AM, AT, AU, AZ, BA, BB, BG, BR, BY, BZ, CA, CH, CN, CO, CR, CU, CZ, DE, DK, DM, DZ, EC, EE, ES, FI, GB, GD, GE, GH, GM, HR, HU, ID, IL, IN, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MA, MD, MG, MK, MN, MW, MX, MZ, NO, NZ, PH, PL, PT, RO, RU, SD, SE, SG, SI, SK, SL, TJ, TM, TR, TT, TZ, UA, UG, UZ, VN, YU, ZA, ZW.
- (84) Designated States (*regional*): ARIPO patent (GH, GM, KE, LS, MW, MZ, SD, SL, SZ, TZ, UG, ZW), Eurasian patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European patent (AT, BE, CH, CY, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE, TR), OAPI patent (BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW, ML, MR, NE, SN, TD, TG).

Published:

— without international search report and to be republished upon receipt of that report

For two-letter codes and other abbreviations, refer to the "Guidance Notes on Codes and Abbreviations" appearing at the beginning of each regular issue of the PCT Gazette.

WO 02/42439 A2

(54) Title: AGGRECANASE MOLECULES

(57) Abstract: Aggrecanase proteins and the nucleotides sequences encoding them as well as processes for producing them are disclosed. Methods for developing inhibitors of the aggrecanase enzymes and antibodies to the enzymes for treatment of conditions characterized by the degradation of aggrecan are also disclosed.

TITLE OF THE INVENTION

AGGRECANASE MOLECULES

The present invention relates to the discovery of nucleotide sequences encoding novel aggrecanase molecules, the aggrecanase proteins and processes for producing them.

5 The invention further relates to the development of inhibitors of, as well as antibodies to the aggrecanase enzymes. These inhibitors and antibodies may be useful for the treatment of various aggrecanase-associated conditions including osteoarthritis.

BACKGROUND OF THE INVENTION

Aggrecan is a major extracellular component of articular cartilage. It is a
10 proteoglycan responsible for providing cartilage with its mechanical properties of compressibility and elasticity. The loss of aggrecan has been implicated in the degradation of articular cartilage in arthritic diseases. Osteoarthritis is a debilitating disease which affects at least 30 million Americans [MacLean et al. J Rheumatol 25:2213-8. (1998)]. Osteoarthritis can severely reduce quality of life due to degradation
15 of articular cartilage and the resulting chronic pain. An early and important characteristic of the osteoarthritic process is loss of aggrecan from the extracellular matrix [Brandt, KD. and Mankin HJ. Pathogenesis of Osteoarthritis, in Textbook of Rheumatology, WB Saunders Company, Philadelphia, PA pgs. 1355-1373. (1993)]. The large, sugar-

containing portion of aggrecan is thereby lost from the extra-cellular matrix, resulting in deficiencies in the biomechanical characteristics of the cartilage.

A proteolytic activity termed "aggrecanase" is thought to be responsible for the cleavage of aggrecan thereby having a role in cartilage degradation associated with osteoarthritis and inflammatory joint disease. Work has been conducted to identify the enzyme responsible for the degradation of aggrecan in human osteoarthritic cartilage. Two enzymatic cleavage sites have been identified within the interglobular domain of aggrecan. One (Asn³⁴¹-Phe³⁴²) is observed to be cleaved by several known metalloproteases [Flannery, CR et al. J Biol Chem 267:1008-14. 1992; Fosang, AJ et al. Biochemical J. 304:347-351. (1994)]. The aggrecan fragment found in human synovial fluid, and generated by IL-1 induced cartilage aggrecan cleavage is at the Glu³⁷³-Ala³⁷⁴ bond [Sandy, JD, et al. J Clin Invest 69:1512-1516. (1992); Lohmander LS, et al. Arthritis Rheum 36: 1214-1222. (1993); Sandy JD et al. J Biol Chem. 266: 8683-8685. (1991)], indicating that none of the known enzymes are responsible for aggrecan cleavage in vivo.

Recently, identification of two enzymes, aggrecanase-1(ADAMTS 4) and aggrecanase -2 (ADAMTS-11) within the "Disintegrin-like and Metalloprotease with Thrombospondin type 1 motif" (ADAM-TS) family have been identified which are synthesized by IL-1 stimulated cartilage and cleave aggrecan at the appropriate site [Tortorella MD, et al Science 284:1664-6. (1999); Abbaszade, I, et al. J Biol Chem 274: 23443-23450. (1999)]. It is possible that these enzymes could be synthesized by osteoarthritic human articular cartilage. It is also contemplated that there are other,

related enzymes in the ADAM-TS family which are capable of cleaving aggrecan at the Glu³⁷³-Ala³⁷⁴ bond and could contribute to aggrecan cleavage in osteoarthritis.

SUMMARY OF THE INVENTION

The present invention is directed to the identification of aggrecanase protein
5 molecules capable of cleaving aggrecanase, the nucleotide sequences which encode the
aggrecanase enzymes, and processes for the production of aggrecanases. These enzymes
are contemplated to be characterized as having proteolytic aggrecanase activity. The
invention further includes compositions comprising these enzymes as well as antibodies
to these enzymes. In addition, the invention includes methods for developing inhibitors
10 of aggrecanase which block the enzyme's proteolytic activity. These inhibitors and
antibodies may be used in various assays and therapies for treatment of conditions
characterized by the degradation of articular cartilage.

The nucleotide sequence of the aggrecanase molecule of the present invention is
set forth in SEQ ID NO:8. In a further embodiment, the nucleotide sequence of the
15 aggrecanase molecule of the present invention is set forth SEQ ID NO: 6 from nucleotide
1 to #5605. Other embodiments of the nucleotide sequence of the invention comprise
the sequences of SEQ ID NO: 1, SEQ ID NO. 2, SEQ ID NO: 3, SEQ ID NO 4 and SEQ
ID NO: 5. The invention further includes equivalent degenerative codon sequences of
these nucleotide sequences, as well as fragments thereof which exhibit aggrecanase
20 activity.

The amino acid sequence of an isolated aggrecanase molecule of the present invention is set forth in SEQ ID NO:9. In a further embodiment, the amino acid sequence of an isolated aggrecanase molecule comprises the sequence set forth in SEQ ID. No. 7.

The invention further includes fragments of the amino acid sequence which encode molecules exhibiting aggrecanase activity. In another embodiment the amino acid sequences of an isolated aggrecanase molecule of the present invention comprises the sequence set forth in SEQ ID NO: 9 or SEQ ID NO: 7 from amino acid #1 to #139.

The human aggrecanase protein or a fragment thereof may be produced by culturing a cell transformed with a DNA sequence of SEQ ID NO: 8 or SEQ ID NO: 6 comprising nucleotide # 1 to #5605 and recovering and purifying from the culture medium a protein characterized by the amino acid sequence set forth in SEQ ID NO: 9 or SEQ ID NO: 7, respectively, substantially free from other proteinaceous materials with which it is co-produced. In another embodiment the human aggrecanase protein or a fragment thereof may be produced by culturing a cell transformed with a DNA sequence of SEQ ID NO: 8 or SEQ ID NO: 6 comprising nucleotide # 1 to #466 of SEQ ID NO:6 and recovering and purifying from the culture medium a protein characterized by the amino acid sequence set forth respectively in SEQ ID NO:9 or SEQ ID NO: 7 comprising amino acid #1-139 substantially free from other proteinaceous materials with which it is co-produced. For production in mammalian cells, the DNA sequence further comprises a DNA sequence encoding a suitable propeptide 5' to and linked in frame to the nucleotide sequence encoding the aggrecanase enzyme.

The invention includes methods for obtaining the full length aggrecanase molecules, the DNA sequence obtained by this method and the protein encoded thereby. The method for isolation of further sequence involves utilizing the aggrecanase sequence set forth in SEQ ID NO:8 or SEQ ID NO: 6 from nucleotide # 1 to #5605 to design
5 probes for screening using standard procedures known to those skilled in the art.

It is expected that other species have DNA sequences homologous to human aggrecanase enzyme. The invention, therefore, includes methods for obtaining the DNA sequences encoding other aggrecanase molecules , the DNA sequences obtained by those methods, and the protein encoded by those DNA sequences. This method entails
10 utilizing the nucleotide sequence of the invention or portions thereof to design probes to screen libraries for the corresponding gene from other species or coding sequences or fragments thereof from using standard techniques. Thus, the present invention may include DNA sequences from other species, which are homologous to the human aggrecanase protein and can be obtained using the human sequence. The present
15 invention may also include functional fragments of the aggrecanase protein, and DNA sequences encoding such functional fragments, as well as functional fragments of other related proteins. The ability of such a fragment to function is determinable by assay of the protein in the biological assays described for the assay of the aggrecanase protein.

The aggrecanase proteins of the present invention may be produced by culturing
20 a cell transformed with the DNA sequence of SEQ ID NO: 8 or the sequence of SEQ ID NO. 6 comprising nucleotide # 1 to # 5605 or comprising nucleotide # 1 to #466 of SEQ ID NO: 6 and recovering and purifying aggrecanase protein from the culture medium. In

the first embodiment the protein comprises the amino acid sequence of SEQ ID NO:9 . In the latter embodiments the protein comprises respectively, amino acid #1 to #1610 of SEQ ID NO:7 and amino acid #1 to #139 of SEQ ID No:7. In further embodiments the nucleotide sequences set forth in SEQ ID NOS: 1, 2, 3, 4, and 5 are utilized in the

5 expression of the aggrecanase molecules. The purified expressed protein is substantially free from other proteinaceous materials with which it is co-produced, as well as from other contaminants. The recovered purified protein is contemplated to exhibit proteolytic aggrecanase activity cleaving aggrecan. Thus, the proteins of the invention may be further characterized by the ability to demonstrate aggrecan proteolytic activity in an

10 assay which determines the presence of an aggrecan-degrading molecule. These assays or the development thereof is within the knowledge of one skilled in the art. Such assays may involve contacting an aggrecan substrate with the aggrecanase molecule and monitoring the production of aggrecan fragments [see for example, Hughes et al., Biochem J 305: 799-804(1995); Mercuri et al, J. Bio Chem. 274:32387-32395 (1999)]

15 In another embodiment, the invention includes methods for developing inhibitors of aggrecanase and the inhibitors produced thereby. These inhibitors prevent cleavage of aggrecan. The method may entail the determination of binding sites based on the three dimensional structure of aggrecanase and aggrecan and developing a molecule reactive with the binding site. Candidate molecules are assayed for inhibitory activity. Additional

20 standard methods for developing inhibitors of the aggrecanase molecule are known to those skilled in the art. Assays for the inhibitors involve contacting a mixture of aggrecan and the inhibitor with an aggrecanase molecule followed by measurement of the

aggrecanase inhibition, for instance by detection and measurement of aggrecan fragments produced by cleavage at an aggrecanase susceptible site.

Another aspect of the invention therefore provides pharmaceutical compositions containing a therapeutically effective amount of aggrecanase inhibitors, in a
5 pharmaceutically acceptable vehicle. Aggrecanase-mediated degradation of aggrecan in cartilage has been implicated in osteoarthritis and other inflammatory diseases. Therefore, these compositions of the invention may be used in the treatment of diseases characterized by the degradation of aggrecan and/or an upregulation of aggrecanase. The compositions may be used in the treatment of these conditions or in the prevention
10 thereof.

The invention includes methods for treating patients suffering from conditions characterized by a degradation of aggrecan or preventing such conditions. These methods, according to the invention, entail administering to a patient needing such treatment, an effective amount of a composition comprising an aggrecanase inhibitor
15 which inhibits the proteolytic activity of aggrecanase enzymes.

Still a further aspect of the invention are DNA sequences coding for expression of an aggrecanase protein. Such sequences include the sequence of nucleotides in a 5' to 3' direction illustrated in SEQ ID NO: 1 comprising nucleotide # 1 to # 1506 or comprising nucleotide # 1 to #1028 of SEQ ID NO: 2 or comprising nucleotide # 1 to #1254 of SEQ
20 ID. NO:3 or comprising nucleotide #1 to #687 of SEQ ID NO: 4 or comprising nucleotide # 1 to #466 of SEQ ID NO: 5 or comprising nucleotide # 1 to #5605 of SEQ ID NO:6 , the nucleotide sequence of SEQ ID NO: 8 and DNA sequences which, but for the

degeneracy of the genetic code, are identical to the DNA sequence set forth above, and encode an aggrecanase protein. Further included in the present invention are DNA sequences which hybridize under stringent conditions with the DNA sequence of SEQ ID NO: 1, SEQ ID NO: 2, SEQ ID NO:3, SEQ ID NO:4, SEQ ID NO: 5, SEQ ID NO: 6, or 5 SEQ ID NO: 8 and encode a protein having the ability to cleave aggrecan. Preferred DNA sequences include those which hybridize under stringent conditions [see, T. Maniatis et al, Molecular Cloning (A Laboratory Manual), Cold Spring Harbor Laboratory (1982), pages 387 to 389]. It is generally preferred that such DNA sequences encode a polypeptide which is at least about 80% homologous, and more preferably at 10 least about 90% homologous, to the sequence of set forth in SEQ ID NO: 9 or in SEQ ID NO: 7 from amino acid #1 to #139 or amino acid #1 to #1610. Finally, allelic or other variations of the sequence of SEQ ID NO: 7 from nucleotide #1 to #466 or from #1 to # 5605 or the sequence of SEQ ID NO:9, whether such nucleotide changes result in changes in the peptide sequence or not, but where the peptide sequence still has 15 aggrecanase activity, are also included in the present invention. The present invention also includes fragments of the DNA sequences shown in SEQ ID NO: 1, SEQ ID NO: 2, SEQ ID NO:3, SEQ ID NO:4, SEQ ID NO: 5, SEQ ID NO: 6 or SEQ ID NO:8 which encode a polypeptide which retains the activity of aggrecanase.

The DNA sequences of the present invention are useful, for example, as probes for 20 the detection of mRNA encoding aggrecanase in a given cell population. Thus, the present invention includes methods of detecting or diagnosing genetic disorders involving the aggrecanase, or disorders involving cellular, organ or tissue disorders in which

aggrecanase is irregularly transcribed or expressed. The DNA sequences may also be useful for preparing vectors for gene therapy applications as described below.

A further aspect of the invention includes vectors comprising a DNA sequence as described above in operative association with an expression control sequence therefor.

5 These vectors may be employed in a novel process for producing an aggrecanase protein of the invention in which a cell line transformed with a DNA sequence encoding an aggrecanase protein in operative association with an expression control sequence therefor, is cultured in a suitable culture medium and an aggrecanase protein is recovered and purified therefrom. This process may employ a number of known cells both
10 prokaryotic and eukaryotic as host cells for expression of the polypeptide. The vectors may be used in gene therapy applications. In such use, the vectors may be transfected into the cells of a patient *ex vivo*, and the cells may be reintroduced into a patient. Alternatively, the vectors may be introduced into a patient *in vivo* through targeted transfection.

15 Still a further aspect of the invention are aggrecanase proteins or polypeptides. Such polypeptides are characterized by having an amino acid sequence including the sequence illustrated in SEQ ID NO. 7 comprising amino acid #1 to #139 or amino acids #1 to #1610, the sequence of SEQ ID NO:9 or variants of the amino acid sequences of SEQ ID NO.7 or SEQ ID NO:9, including naturally occurring allelic variants, and other
20 variants in which the protein retains the ability to cleave aggrecan characteristic of aggrecanase molecules. Preferred polypeptides include a polypeptide which is at least about 80% homologous, and more preferably at least about 90% homologous, to the

amino acid sequence shown in SEQ ID NO. 7 comprising amino acid #1 to #139 or
comprising #1 to #1610 or the sequence of SEQ ID NO: 9. Finally, allelic or other
variations of these sequences of SEQ ID NO. 7 or SEQ ID NO:9 , whether such amino
acid changes are induced by mutagenesis, chemical alteration, or by alteration of DNA
5 sequence used to produce the polypeptide, where the peptide sequence still has
aggrecanase activity, are also included in the present invention. The present invention
also includes fragments of the amino acid sequence of SEQ ID NO. 7 or SEQ ID NO:9
which retain the activity of aggrecanase protein.

The purified proteins of the present inventions may be used to generate antibodies,
10 either monoclonal or polyclonal, to aggrecanase and/or other aggrecanase -related
proteins, using methods that are known in the art of antibody production. Thus, the
present invention also includes antibodies to aggrecanase or other related proteins. The
antibodies may be useful for detection and/or purification of aggrecanase or related
proteins, or for inhibiting or preventing the effects of aggrecanase. The aggrecanase of
15 the invention or portions thereof may be utilized to prepare antibodies that specifically
bind to aggrecanase.

DETAILED DESCRIPTION OF THE INVENTION

The nucleotide sequence of the human aggrecanase of the present invention
comprises the sequence set forth in SEQ ID NO:8. In a further embodiment, the
20 nucleotide sequence of the human aggrecanase of the present invention comprises
nucleotides # 1 to # 5605 of SEQ ID NO: 6. In another embodiment the nucleotide

sequence comprises nucleotide #1-466 of SEQ ID NO:6. Other embodiments comprise the nucleotide sequences set forth in SEQ ID NO: 1, SEQ ID NO: 2, SEQ ID NO:3, SEQ ID NO:4, and SEQ ID NO: 5. The human aggrecanase protein sequence comprises the sequence set forth in SEQ ID NO:9. In a further embodiment, a human aggrecanase of
5 the present invention comprises amino acids # 1 to # 1610 set forth in SEQ ID NO. 7. In another embodiment the human aggrecanase sequence of the invention comprises amino acids #1 to #466 of SEQ ID NO: 7. Further sequences of the aggrecanase of the present invention may be obtained using the sequences of SEQ ID NO. 6 comprising nucleotides # 1 to # 466 or nucleotides #1 to # 5605 to design probes for screening for the full
10 sequence using standard techniques.

The aggrecanase proteins of the present invention, include polypeptides comprising the amino acid sequence SEQ ID NO:9 or of SEQ ID NO.7 from amino acid #1 to #139 or from #1 to #1610 and having the ability to cleave aggrecan.

The aggrecanase proteins recovered from the culture medium are purified by
15 isolating them from other proteinaceous materials from which they are co-produced and from other contaminants present. The isolated and purified proteins may be characterized by the ability to cleave aggrecan substrate. The aggrecanase proteins provided herein also include factors encoded by the sequences similar to those of SEQ ID NO: 1, SEQ ID NO: 2, SEQ ID NO:3, SEQ ID NO:4, SEQ ID NO: 5, SEQ ID NO: 6 or
20 SEQ ID NO: 8, but into which modifications or deletions are naturally provided (e.g. allelic variations in the nucleotide sequence which may result in amino acid changes in the polypeptide) or deliberately engineered. For example, synthetic polypeptides may

wholly or partially duplicate continuous sequences of the amino acid residues of SEQ ID NO. 7 or SEQ ID NO: 9. These sequences, by virtue of sharing primary, secondary, or tertiary structural and conformational characteristics with aggrecanase molecules may possess biological properties in common therewith. It is known, for example that

5 numerous conservative amino acid substitutions are possible without significantly modifying the structure and conformation of a protein, thus maintaining the biological properties as well. For example, it is recognized that conservative amino acid substitutions may be made among amino acids with basic side chains, such as lysine (Lys or K), arginine (Arg or R) and histidine (His or H); amino acids with acidic side chains,

10 such as aspartic acid (Asp or D) and glutamic acid (Glu or E); amino acids with uncharged polar side chains, such as asparagine (Asn or N), glutamine (Gln or Q), serine (Ser or S), threonine (Thr or T), and tyrosine (Tyr or Y); and amino acids with nonpolar side chains, such as alanine (Ala or A), glycine (Gly or G), valine (Val or V), leucine (Leu or L), isoleucine (Ile or I), proline (Pro or P), phenylalanine (Phe or F), methionine (Met

15 or M), tryptophan (Trp or W) and cysteine (Cys or C). Thus, these modifications and deletions of the native aggrecanase may be employed as biologically active substitutes for naturally-occurring aggrecanase and in the development of inhibitors other polypeptides in therapeutic processes. It can be readily determined whether a given variant of aggrecanase maintains the biological activity of aggrecanase by subjecting both

20 aggrecanase and the variant of aggrecanase, as well as inhibitors thereof, to the assays described in the examples.

Other specific mutations of the sequences of aggrecanase proteins described herein involve modifications of glycosylation sites. These modifications may involve O-linked or N-linked glycosylation sites. For instance, the absence of glycosylation or only partial glycosylation results from amino acid substitution or deletion at asparagine-linked glycosylation recognition sites. The asparagine-linked glycosylation recognition sites comprise tripeptide sequences which are specifically recognized by appropriate cellular glycosylation enzymes. These tripeptide sequences are either asparagine-X-threonine or asparagine-X-serine, where X is usually any amino acid. A variety of amino acid substitutions or deletions at one or both of the first or third amino acid positions of a glycosylation recognition site (and/or amino acid deletion at the second position) results in non-glycosylation at the modified tripeptide sequence. Additionally, bacterial expression of aggrecanase-related protein will also result in production of a non-glycosylated protein, even if the glycosylation sites are left unmodified.

The present invention also encompasses the novel DNA sequences, free of association with DNA sequences encoding other proteinaceous materials, and coding for expression of aggrecanase proteins. These DNA sequences include those depicted in SEQ ID NO: 8 or SEQ ID NO: 1 in a 5' to 3' direction and those sequences which hybridize thereto under stringent hybridization washing conditions [for example, 0.1X SSC, 0.1% SDS at 65°C; see, T. Maniatis et al, Molecular Cloning (A Laboratory Manual), Cold Spring Harbor Laboratory (1982), pages 387 to 389] and encode a protein having aggrecanase proteolytic activity. These DNA sequences also include those which comprise the DNA sequence of SEQ ID NO: 1, SEQ ID NO: 2, SEQ ID NO:3, SEQ ID

NO:4, SEQ ID NO: 5, SEQ ID NO: 6 and those which hybridize thereto under stringent hybridization conditions and encode a protein which maintain the other activities disclosed for aggrecanase.

Similarly, DNA sequences which code for aggrecanase proteins coded for by the sequences of SEQ ID NO: 1, SEQ ID NO: 2, SEQ ID NO:3, SEQ ID NO:4, SEQ ID NO: 5, or SEQ ID NO: 6 comprising nucleotide # 1 to # 466 or comprising nucleotide # 1 to #5605 of SEQ ID NO 6 or SEQ ID NO: 8 or aggrecanase proteins which comprise the amino acid sequence of SEQ ID NO: 9 or SEQ ID NO.7 from amino acid # 1-139 or #1 to #1610, but which differ in codon sequence due to the degeneracies of the genetic code or allelic variations (naturally-occurring base changes in the species population which may or may not result in an amino acid change) also encode the novel factors described herein. Variations in the DNA sequences of SEQ ID NO: 1, SEQ ID NO: 2, SEQ ID NO:3, SEQ ID NO:4, SEQ ID NO: 5, SEQ ID NO: 6 or SEQ ID NO: 8 comprising nucleotide # 1 to # 466 or comprising nucleotide # 1 to #5605 of SEQ ID NO:7 which are caused by point mutations or by induced modifications (including insertion, deletion, and substitution) to enhance the activity, half-life or production of the polypeptides encoded are also encompassed in the invention.

Another aspect of the present invention provides a novel method for producing aggrecanase proteins. The method of the present invention involves culturing a suitable cell line, which has been transformed with a DNA sequence encoding a aggrecanase protein of the invention, under the control of known regulatory sequences. The transformed host cells are cultured and the aggrecanase proteins recovered and purified

from the culture medium. The purified proteins are substantially free from other proteins with which they are co-produced as well as from other contaminants.

Suitable cells or cell lines may be mammalian cells, such as Chinese hamster ovary cells (CHO). The selection of suitable mammalian host cells and methods for transformation, culture, amplification, screening, product production and purification are known in the art. See, e.g., Gething and Sambrook, Nature, 293:620-625 (1981), or alternatively, Kaufman et al, Mol. Cell. Biol., 5(7):1750-1759 (1985) or Howley et al, U.S. Patent 4,419,446. Another suitable mammalian cell line, which is described in the accompanying examples, is the monkey COS-1 cell line. The mammalian cell CV-1 may also be suitable.

Bacterial cells may also be suitable hosts. For example, the various strains of E. coli (e.g., HB101, MC1061) are well-known as host cells in the field of biotechnology. Various strains of B. subtilis, Pseudomonas, other bacilli and the like may also be employed in this method. For expression of the protein in bacterial cells, DNA encoding the propeptide of Aggrecanase is generally not necessary.

Many strains of yeast cells known to those skilled in the art may also be available as host cells for expression of the polypeptides of the present invention. Additionally, where desired, insect cells may be utilized as host cells in the method of the present invention. See, e.g. Miller et al, Genetic Engineering, 8:277-298 (Plenum Press 1986) and references cited therein.

Another aspect of the present invention provides vectors for use in the method of expression of these novel aggrecanase polypeptides. Preferably the vectors contain the

full novel DNA sequences described above which encode the novel factors of the invention. Additionally, the vectors contain appropriate expression control sequences permitting expression of the aggrecanase protein sequences. Alternatively, vectors incorporating modified sequences as described above are also embodiments of the present
5 invention. Additionally, the sequence of SEQ ID NO: 1, SEQ ID NO: 2, SEQ ID NO:3, SEQ ID NO:4, SEQ ID NO: 5, SEQ ID NO: 6 or SEQ ID NO:8 or other sequences encoding aggrecanase proteins could be manipulated to express composite aggrecanase molecules. Thus, the present invention includes chimeric DNA molecules encoding an aggrecanase protein comprising a fragment from SEQ ID NO:8 or SEQ ID NO: 6
10 comprising nucleotide # 1 to # 466 or comprising nucleotide # 1 to #5605 of SEQ ID NO: 6 linked in correct reading frame to a DNA sequence encoding another aggrecanase polypeptide.

The vectors may be employed in the method of transforming cell lines and contain selected regulatory sequences in operative association with the DNA coding sequences of
15 the invention which are capable of directing the replication and expression thereof in selected host cells. Regulatory sequences for such vectors are known to those skilled in the art and may be selected depending upon the host cells. Such selection is routine and does not form part of the present invention.

Various conditions such as osteoarthritis are known to be characterized by
20 degradation of aggrecan. Therefore, an aggrecanase protein of the present invention which cleaves aggrecan may be useful for the development of inhibitors of aggrecanase. The invention therefore provides compositions comprising an aggrecanase inhibitor. The

inhibitors may be developed using the aggrecanase in screening assays involving a mixture of aggrecan substrate with the inhibitor followed by exposure to aggrecan. The compositions may be used in the treatment of osteoarthritis and other conditions exhibiting degradation of aggrecan.

5 The invention further includes antibodies which can be used to detect aggrecanase and also may be used to inhibit the proteolytic activity of aggrecanase.

 The therapeutic methods of the invention includes administering the aggrecanase inhibitor compositions topically, systemically, or locally as an implant or device. The dosage regimen will be determined by the attending physician considering various factors
10 which modify the action of the aggrecanase protein, the site of pathology, the severity of disease, the patient's age, sex, and diet, the severity of any inflammation, time of administration and other clinical factors. Generally, systemic or injectable administration will be initiated at a dose which is minimally effective, and the dose will be increased over a preselected time course until a positive effect is observed. Subsequently,
15 incremental increases in dosage will be made limiting such incremental increases to such levels that produce a corresponding increase in effect, while taking into account any adverse affects that may appear. The addition of other known factors, to the final composition, may also effect the dosage.

 Progress can be monitored by periodic assessment of disease progression. The
20 progress can be monitored, for example, by x-rays, MRI or other imaging modalities, synovial fluid analysis, and/or clinical examination.

The following examples illustrate practice of the present invention in isolating and characterizing human aggrecanase and other aggrecanase-related proteins, obtaining the human proteins and expressing the proteins via recombinant techniques.

EXAMPLES

5 **EXAMPLE 1**

Isolation of DNA

Potential novel aggrecanase family members were identified using a database screening approach. Aggrecanase-1 [Science284:1664-1666 (1999)] has at least six domains: signal, propeptide, catalytic domain, disintegrin, tsp and c-terminal. The catalytic
10 domain contains a zinc binding signature region, TAAHELGHVKF and a "MET turn" which are responsible for protease activity. Substitutions within the zinc binding region in the number of the positions still allow protease activity, but the histidine (H) and glutamic acid (E) residues must be present. The thrombospondin domain of Aggrecanase-1 is also a critical domain for substrate recognition and cleavage. It is these
15 two domains that determine our classification of a novel aggrecanase family member. The coding region of the Aggrecanase-1 DNA sequence was used to query against the GeneBank ESTs focusing on human ESTs using TBLASTN. The resulting sequences were the starting point in the effort to identify full length sequence for potential family members. The nucleotide sequence of the aggrecanase of the present invention is
20 comprised of one EST (A1479925) that contains homology over the catalytic domain and zinc binding motif of Aggrecanase-1(ADAMTS4).

AI479925 was used to search the public database using the algorithm BLASTX, which searches a protein sequence database using all six conceptual translations of a nucleotide sequence query. AI479925 was shown to have 98% homology to KIAA1312 over 83 bps. The KIAA1312 sequence was used to query the public databases with the algorithm BLASTX and found to have 44% identity to ADAMTS-1. KIAA1312 was sequenced by the Kazusa DNA Research Institute. KIAA1312 appears to be 5' truncated missing the signal and propeptide. KIAA1312 contains the catalytic domain, disintegrin, tsp type I motif and c-terminal spacer(found in ADAMTS4. It is with these criteria that candidate #7 (KIAA1312) is considered a novel Aggrecanase family member.

10 GenBank deposit (for ADAMTS9) showed identity to EST7 and KIAA1312. By alignment with other family members, ADAMTS9 appears to have intact 5P signal and propetide sequences, but is 3P truncated in comparison to KIAA1312. A full-length EST7 sequence has been constructed using the initiator met from ADAMTS9 and the translational stop found in KIAA1312.

15 This human aggrecanase sequences were isolated from a dT-primed cDNA library constructed in the plasmid vector pED6-dpc2. cDNA was made from human small intestine RNA purchased from Clontech. The probe to isolate the aggrecanase of the present invention was generated from the sequence obtained from the database search. The sequence of the probe was as follows:

20 EST7_1,_4,_8,_9

5'-GACAGCTTTTACGATCGCCCATGAGCTG-3'

EST7_8B

5'-TTAATGCTGCTACGTCACCAGCCAGGTTA-3'

The DNA probe was radioactively labeled with ³²P and used to screen the human small intestine dT-primed cDNA library, under high stringency hybridization/washing conditions, to identify clones containing sequences of the human candidate #7.

Nitrocellulose replicas of the transformed colonies were hybridized to the ³²P labeled DNA probe in standard hybridization buffer (1X Blotto[25X Blotto = %5 nonfat dried milk, 0.02% azide in dH₂O] + 1% NP-40 + 6X SSC + 0.05% Pyrophosphate) under high stringency conditions (65°C for 2 hours). After 2 hours hybridization, the radioactively labeled DNA probe containing hybridization solution was removed and the filters were washed under high stringency conditions (3X SSC, 0.05% Pyrophosphate for 5 minutes at RT; followed by 2.2X SSC, 0.05% Pyrophosphate for 15 minutes at RT; followed by 2.2X SSC, 0.05% Pyrophosphate for 1-2 minutes shaking at 65°C. The filters were wrapped in Saran wrap and exposed to X-ray film for overnight. The autoradiographs were developed and positively hybridizing transformants of various signal intensities were identified. These positive clones were picked; grown for 12 hours in selective medium and plated at low density (approximately 100 colonies per plate). Nitrocellulose replicas of the colonies were hybridized to the ³²P labeled probe in standard hybridization buffer ((1X Blotto[25X Blotto = %5 nonfat dried milk, 0.02% azide in dH₂O] + 1% NP-40 + 6X SSC + 0.05% Pyrophosphate) under high stringency conditions (65°C for 2 hours). After 2 hours hybridization, the radioactively labeled DNA probe containing hybridization solution was removed and the filters were washed

under high stringency conditions (3X SSC, 0.05% Pyrophosphate for 5 minutes at RT standing; followed by 2.2X SSC, 0.05% Pyrophosphate for 15 minutes shaking at RT; followed by 2.2X SSC, 0.05% Pyrophosphate for 1-2 minutes shaking at 65°C. The filters were wrapped in Saran wrap and exposed to X-ray film for overnight. The

5 autoradiographs were developed and positively hybridizing transformants were identified. Bacterial stocks of purified hybridization positive clones were made and plasmid DNA was isolated. The sequences of the cDNA inserts were determined and are set forth in SEQ ID NO: 1, SEQ ID NO: 2, SEQ ID NO:3, SEQ ID NO:4, and SEQ ID NO: 5. Sequences have been deposited in the American Type Culture Collection 10801

10 University Blvd. Manassas, VA 20110-2209 USA as PTA -2283. The cDNA insert contained the sequences of the DNA probe used in the hybridization.

SEQ ID NO: 1 sets forth EST7.1 comprising nucleotides #1 to #1506. SEQ ID NO: 1 aligns with KIAA1312 from 195 to 330 amino acids. Nucleotides #1081 to #1245 of SEQ ID NO: 1 correspond with nucleotides #1161 to #1456 of SEQ ID NO:6.

15 SEQ ID NO: 2 sets forth EST 7.4 comprising nucleotides #1 to #1028. SEQ ID NO: 2 aligns with KIAA1312 from 47 to 330 amino acids. Nucleotides # 22 to #872 of SEQ ID NO 2 corresponds with nucleotides #606 to #1456 of SEQ ID NO:6.

SEQ ID NO: 3 sets forth EST 7.8 comprising nucleotides #1 - #12054. SEQ ID NO: 3 aligns with KIAA1312 from 195 to 281 amino acids. SEQ ID NO: 4 sets forth EST 7.9

20 comprising nucleotides #1 to #687. EST 7.9 aligns with KIAA1312 from 149-330 amino acids. SEQ ID NO: 4 nucleotides #14 to #555 correspond with nucleotides # 915 to #1456 of SEQ ID NO: 6 .SEQ ID NO: 5 sets forth EST 7.8B comprising nucleotides #1

to #466 which extends the 5' sequence of KIAA1312. SEQ ID NO: 6 comprising nucleotides #1 to # 5605 sets forth the nucleotide sequence of SEQ ID NO: 5 (nucleotides #1-#466) and KIAA1312 (nucleotides # 467 to # 5605). SEQ ID NO:7 comprising amino acids #1 to #1610 sets forth the amino acid sequence of EST

5 7.8B (amino acids #1-#139) and the amino acids of KIAA1312 (amino acids #140 to #1610).

This human Aggrecanase gene, full-length EST7, was isolated using a PCR strategy. Tissue sources were identified by a PCR screen of various phage and plasmid libraries using oligos designed to EST7. EST7 was found expressed in small intestine,
 10 brain and kidney libraries. Based on the publicly available sequence, PCR primers to the full-length EST7 sequence were designed. Three overlapping pieces of EST7 were amplified using the following primer sets. The first PCR primer set amplified from bp 1-1864 of the full-length EST7 sequence; 5P primer sequence –
 ATAAGATTGCGGCCGCCACCATGCAGTTTGTATCCTGGGCCACAC (this primer
 15 incorporated an 8 bp tail (ATAAGATT), a NOT1 sequence (GCGGCCGC) and a Kozak sequence (CCACC) upstream of the initiator Met (ATG)) and the 3P primer sequence –
 GTCTGTTGCACTCTCGAATGGCTGTT. The second primer set amplified from bp 1752-3485 of the full-length EST7 sequence; 5P primer sequence –
 AGAAATGGATGTCCCCGTGACAGATG and the 3P primer sequence –
 20 TGGGTATCAGTTGGTCTAGTTGCTGC. The third PCR primer set amplified from bp 3300-5054 of the full-length EST7 sequence; 5P primer sequence -
 GCCAACATCTATGCAGACTTGTCAGC and 3P primer sequence –

GGAATTCTAGCTTGGGAAAGCTGAGGA. The Advantage-GC 2 PCR Kit from Clontech was used to amplify the full-length EST7 gene products. Reaction conditions were those recommended in the user manual; with the following exceptions: per 50 ul reaction the amount of GC Melt used was 5 ul; the amount of phage library (Clontech human kidney 5'-STRETCH PLUS cDNA library) used was 2 ul of a stock with titer $\geq 10^8$ pfu/ml or 10ng plasmid library DNA linearized with Not1, and the amount of each PCR primer used was 1ul of a 10 pmol/ul stock. Cycling conditions were as follows: 94°C for 1 min, one cycle; followed by 40 cycles consisting of 94°C for 15 sec/68°C for 3 min. The primer pairs were used in PCR amplification reactions containing each of the 3 tissue sources; kidney, small intestine or brain. PCR products resulting from the amplification of the 5P 1864 base pair product were digested with Not1 and BamH1 and ligated into the CS2+ vector (digested with the same) using standard digestion and ligation conditions. PCR products resulting from the amplification of the internal 1733 base pair product were digested with BamH1 and Nsi1 and ligated into the CS2+ vector (digested with the same) using standard digestion and ligation conditions. PCR products resulting from the amplification of the 3P 1754 base pair product were digested with Nsi1 and EcoR1 and ligated into the CS2+ vector (digested with the same) using standard digestion and ligation conditions. Ligated products were transformed into ElectroMAX DH10B cells from Life Technologies. Cloned PCR fragments of EST7 were sequenced to determine fidelity. The full-length sequence for EST7 was the consensus sequence derived from the KIAA1312 and the ADAMTS9 sequences. PCR products with the correct sequence were excised from the CS2+ vector using the appropriate enzyme pairs

described above. A full-length version of EST7 was constructed by ligating these 3 PCR products, 5P (Not1/BamH1), internal (BamH1/Nsi1) and 3P (Nsi1/EcoR1), into the Cos expression vector pED6-dpc1 (digested with Not1 and EcoR1).

The full length ADAMTS9 EST7 sequence in pED6-dpc1:Not 1 to EcoR1 is set
5 forth in SEQ ID NO:8. The peptide sequence is set forth in SEQ ID NO:9.

EXAMPLE 2

Expression of Aggrecanase

In order to produce murine, human or other mammalian aggrecanase-related proteins, the DNA encoding it is transferred into an appropriate expression vector and
10 introduced into mammalian cells or other preferred eukaryotic or prokaryotic hosts including insect host cell culture systems by conventional genetic engineering techniques. Expression system for biologically active recombinant human aggrecanase is contemplated to be stably transformed mammalian cells, insect, yeast or bacterial cells.

One skilled in the art can construct mammalian expression vectors by employing
15 the sequence of SEQ ID NO; 8 or SEQ ID NO: 6 comprising nucleotide # 1 to # 466 or comprising nucleotide # 1 to #5605 of SEQ ID NO 6 or the DNA sequences of SEQ ID NO: 1, SEQ ID NO: 2, SEQ ID NO:3, SEQ ID NO:4, or SEQ ID NO: 5 encoding aggrecanase-related proteins or other modified sequences and known vectors, such as pCD [Okayama et al., Mol. Cell Biol., 2:161-170 (1982)], pJL3, pJL4 [Gough et al.,
20 EMBO J., 4:645-653 (1985)] and pMT2 CXM.

The mammalian expression vector pMT2 CXM is a derivative of p91023(b) (Wong et al., Science 228:810-815, 1985) differing from the latter in that it contains the ampicillin resistance gene in place of the tetracycline resistance gene and further contains a XhoI site for insertion of cDNA clones. The functional elements of pMT2 CXM have
5 been described (Kaufman, R.J., 1985, Proc. Natl. Acad. Sci. USA 82:689-693) and include the adenovirus VA genes, the SV40 origin of replication including the 72 bp enhancer, the adenovirus major late promoter including a 5' splice site and the majority of the adenovirus tripartite leader sequence present on adenovirus late mRNAs, a 3' splice acceptor site, a DHFR insert, the SV40 early polyadenylation site (SV40), and pBR322
10 sequences needed for propagation in E. coli.

Plasmid pMT2 CXM is obtained by EcoRI digestion of pMT2-VWF, which has been deposited with the American Type Culture Collection (ATCC), Rockville, MD (USA) under accession number ATCC 67122. EcoRI digestion excises the cDNA insert present in pMT2-VWF, yielding pMT2 in linear form which can be ligated and used to
15 transform E. coli HB 101 or DH-5 to ampicillin resistance. Plasmid pMT2 DNA can be prepared by conventional methods. pMT2 CXM is then constructed using loopout/in mutagenesis [Morinaga, et al., Biotechnology 84: 636 (1984). This removes bases 1075 to 1145 relative to the Hind III site near the SV40 origin of replication and enhancer sequences of pMT2. In addition it inserts the following sequence:

20 5' PO-CATGGGCAGCTCGAG-3'

at nucleotide 1145. This sequence contains the recognition site for the restriction endonuclease Xho I. A derivative of pMT2CXM, termed pMT23, contains recognition

sites for the restriction endonucleases PstI, Eco RI, SalI and XhoI. Plasmid pMT2 CXM and pMT23 DNA may be prepared by conventional methods.

pEMC2 β 1 derived from pMT21 may also be suitable in practice of the invention. pMT21 is derived from pMT2 which is derived from pMT2-VWF. As described above
 5 EcoRI digestion excises the cDNA insert present in pMT-VWF, yielding pMT2 in linear form which can be ligated and used to transform E. Coli HB 101 or DH-5 to ampicillin resistance. Plasmid pMT2 DNA can be prepared by conventional methods.

pMT21 is derived from pMT2 through the following two modifications. First, 76 bp of the 5' untranslated region of the DHFR cDNA including a stretch of 19 G residues
 10 from G/C tailing for cDNA cloning is deleted. In this process, a XhoI site is inserted to obtain the following sequence immediately upstream from DHFR: 5' -

CTGCAGGCGAGCCTGAATTCCTCGAGCCATCATG-3'

PstI

Eco RI XhoI

Second, a unique ClaI site is introduced by digestion with EcoRV and XbaI, treatment
 15 with Klenow fragment of DNA polymerase I, and ligation to a ClaI linker (CATCGATG). This deletes a 250 bp segment from the adenovirus associated RNA (VAI) region but does not interfere with VAI RNA gene expression or function. pMT21 is digested with EcoRI and XhoI, and used to derive the vector pEMC2B1.

A portion of the EMCV leader is obtained from pMT2-ECAT1 [S.K. Jung, et al, J. Virol
 20 63:1651-1660 (1989)] by digestion with Eco RI and PstI, resulting in a 2752 bp fragment. This fragment is digested with TaqI yielding an Eco RI-TaqI fragment of 508

bp which is purified by electrophoresis on low melting agarose gel. A 68 bp adapter and its complementary strand are synthesized with a 5' TaqI protruding end and a 3' XhoI protruding end which has the following sequence:

5'-CGAGGTTAAAAACGTCTAGGCCCCCCGAACCACGGGGACGTGGTTTTCCTTT

5 TaqI

GAAAAACACGATTGC-3'

 XhoI

This sequence matches the EMC virus leader sequence from nucleotide 763 to 827. It
 10 also changes the ATG at position 10 within the EMC virus leader to an ATT and is followed by a XhoI site. A three way ligation of the pMT21 Eco RI-16hoI fragment, the EMC virus EcoRI-TaqI fragment, and the 68 bp oligonucleotide adapter TaqI-16hoI adapter resulting in the vector pEMC2 β 1:

This vector contains the SV40 origin of replication and enhancer, the adenovirus
 15 major late promoter, a cDNA copy of the majority of the adenovirus tripartite leader sequence, a small hybrid intervening sequence, an SV40 polyadenylation signal and the adenovirus VA I gene, DHFR and β -lactamase markers and an EMC sequence, in appropriate relationships to direct the high level expression of the desired cDNA in mammalian cells.

20 The construction of vectors may involve modification of the aggrecanase-related DNA sequences. For instance, aggrecanase cDNA can be modified by removing the non-

coding nucleotides on the 5' and 3' ends of the coding region. The deleted non-coding nucleotides may or may not be replaced by other sequences known to be beneficial for expression. These vectors are transformed into appropriate host cells for expression of aggrecanase-related proteins. Additionally, the sequence of SEQ ID NO: 8 or the
5 sequence of SEQ ID NO: 6 comprising nucleotide # 1 to # 466 or comprising nucleotide # 1 to #5605 of SEQ ID NO: 6 or other sequences encoding aggrecanase-related proteins such as the sequences comprising SEQ ID NO: 1, SEQ ID NO: 2, SEQ ID NO:3, SEQ ID NO:4, or SEQ ID NO: 5 can be manipulated to express a mature aggrecanase-related protein by deleting aggrecanase encoding propeptide sequences and replacing them with
10 sequences encoding the complete propeptides of other aggrecanase proteins.

One skilled in the art can manipulate the sequences of SEQ ID NO: 1 through SEQ ID NO: 6 or SEQ ID NO:8 by eliminating or replacing the mammalian regulatory sequences flanking the coding sequence with bacterial sequences to create bacterial vectors for intracellular or extracellular expression by bacterial cells. For example, the
15 coding sequences could be further manipulated (e.g. ligated to other known linkers or modified by deleting non-coding sequences therefrom or altering nucleotides therein by other known techniques). The modified aggrecanase-related coding sequence could then be inserted into a known bacterial vector using procedures such as described in T. Taniguchi et al., Proc. Natl Acad. Sci. USA, 77:5230-5233 (1980). This exemplary
20 bacterial vector could then be transformed into bacterial host cells and a aggrecanase-related protein expressed thereby. For a strategy for producing extracellular expression of

aggrecanase-related proteins in bacterial cells, see, e.g. European patent application EPA 177,343.

Similar manipulations can be performed for the construction of an insect vector [See, e.g. procedures described in published European patent application 155,476] for expression in insect cells. A yeast vector could also be constructed employing yeast regulatory sequences for intracellular or extracellular expression of the factors of the present invention by yeast cells. [See, e.g., procedures described in published PCT application WO86/00639 and European patent application EPA 123,289].

A method for producing high levels of a aggrecanase-related protein of the invention in mammalian, bacterial, yeast or insect host cell systems may involve the construction of cells containing multiple copies of the heterologous Aggrecanase-related gene. The heterologous gene is linked to an amplifiable marker, e.g. the dihydrofolate reductase (DHFR) gene for which cells containing increased gene copies can be selected for propagation in increasing concentrations of methotrexate (MTX) according to the procedures of Kaufman and Sharp, J. Mol. Biol., 159:601-629 (1982). This approach can be employed with a number of different cell types.

For example, a plasmid containing a DNA sequence for an aggrecanase-related protein of the invention in operative association with other plasmid sequences enabling expression thereof and the DHFR expression plasmid pAdA26SV(A)3 [Kaufman and Sharp, Mol. Cell. Biol., 2:1304 (1982)] can be co-introduced into DHFR-deficient CHO cells, DUKX-BII, by various methods including calcium phosphate coprecipitation and transfection, electroporation or protoplast fusion. DHFR expressing transformants are

selected for growth in alpha media with dialyzed fetal calf serum, and subsequently selected for amplification by growth in increasing concentrations of MTX (e.g. sequential steps in 0.02, 0.2, 1.0 and 5uM MTX) as described in Kaufman et al., Mol Cell Biol., 5:1750 (1983). Transformants are cloned, and biologically active aggrecanase expression is monitored by the assays described above. Aggrecanase protein expression should increase with increasing levels of MTX resistance. Aggrecanase polypeptides are characterized using standard techniques known in the art such as pulse labeling with [35S] methionine or cysteine and polyacrylamide gel electrophoresis. Similar procedures can be followed to produce other related aggrecanase-related proteins.

10 In one example the aggrecanase gene of the present invention set forth in SEQ ID NO: 8 is cloned into the expression vector pED6 [Kaufman et al., Nucleic Acid Res. 19:44885-4490(1991)]. COS and CHO DUKX B11 cells are transiently transfected with the aggrecanase sequence of the invention (+/- co-transfection of PACE on a separate pED6 plasmid) by lipofection(LF2000, Invitrogen). Duplicate tranfections are performed for each gene of interest: (a) one for harvesting conditioned media for activity assay and (b) one for 35-S-methionine/cysteine metabolic labeling.

On day one media is changed to DME(COS) or alpha(CHO) media + 1% heat-inactivated fetal calf serum +/- 100µg/ml heparin on wells(a) to be harvested for activity assay. After 48h (day4), conditioned media is harvested for activity assay.

20 On day 3, the duplicate wells(b) were changed to MEM (methionine-free/cysteine free) media + 1% heat-inactivated fetal calf serum +100µg/ml heparin + 100µCi/ml 35S-methioine/cysteine (Redivue Pro mix, Amersham). Following 6h incubation at 37°C,

conditioned media is harvested and run on SDS-PAGE gels under reducing conditions. Proteins are visualized by autoradiography.

EXAMPLE 3

5 Biological Activity of Expressed Aggrecanase

To measure the biological activity of the expressed aggrecanase-related proteins obtained in Example 2 above, the proteins are recovered from the cell culture and purified by isolating the aggrecanase-related proteins from other proteinaceous materials with which they are co-produced as well as from other contaminants. The purified protein may
10 be assayed in accordance with assays described above. Purification is carried out using standard techniques known to those skilled in the art.

Protein analysis is conducted using standard techniques such as SDS-PAGE acrylamide [Laemmli, Nature 227:680 (1970)] stained with silver [Oakley, et al. Anal. Biochem. 105:361 (1980)] and by immunoblot [Towbin, et al. Proc. Natl. Acad. Sci. USA
15 76:4350 (1979)]

The foregoing descriptions detail presently preferred embodiments of the present invention. Numerous modifications and variations in practice thereof are expected to occur to those skilled in the art upon consideration of these descriptions. Those modifications and variations are believed to be encompassed within the claims appended
20 hereto.

What is claimed is:

1. An isolated DNA molecule comprising a DNA sequence set forth in SEQ ID NO:8.
2. An isolated DNA molecule comprising a DNA sequence set forth in SEQ ID NO. 6.
3. An isolated DNA molecule comprising a DNA sequence set forth in SEQ ID NO. 5.
4. An isolated DNA molecule comprising a DNA sequence selected from the group consisting of
 - a) the sequence of SEQ ID No. 1,
 - b) the sequence of SEQ ID NO: 2,
 - c) the sequence of SEQ ID NO: 3,
 - d) the sequence of SEQ ID NO: 4,
 - c) naturally occurring human allelic sequences and equivalent degenerative codon sequences of a) through d).
5. A vector comprising a DNA molecule of claim 1 in operative association with an expression control sequence therefor.

6. A vector comprising a DNA molecule of claim 2 in operative association with an expression control sequence therefor.
7. A host cell transformed with the DNA sequence of claim 1.
8. A host cell transformed with a DNA sequence of claim 2.
- 5 9. A method for producing a purified human aggrecanase protein, said method comprising the steps of:
 - (a) culturing a host cell transformed with a DNA molecule according to claim 1; and
 - (b) recovering and purifying said aggrecanase protein comprising the amino
10 acid sequence of SEQ ID NO:9 from the culture medium.
10. A method for producing a purified human aggrecanase protein, said method comprising the steps of:
 - (a) culturing a host cell transformed with a DNA molecule according to claim 2; and
 - 15 (b) recovering and purifying said aggrecanase protein comprising amino acid sequence of SEQ ID NO:7 from the culture medium.

11. A purified aggrecanase polypeptide comprising the amino acid sequence set forth in SEQ ID NO: 9.
12. A purified aggrecanase polypeptide comprising the amino acid sequence set forth in SEQ ID NO 7.
- 5 13. A purified aggrecanase polypeptide produced by the steps of
 - (a) culturing a cell transformed with a DNA molecule according to claim 1; and
 - (b) recovering and purifying from said culture medium a polypeptide comprising the amino acid sequence set forth in SEQ ID NO. 97.
- 10 14. A purified aggrecanase polypeptide produced by the steps of
 - (a) culturing a cell transformed with a DNA molecule according to claim 2; and
 - (b) recovering and purifying from said culture medium a polypeptide comprising the amino acid sequence set forth in SEQ ID NO. 7.
- 15 15. An antibody that binds to a purified aggrecanase protein of claim 11.
- 15 16. An antibody that binds to a purified aggrecanase protein of claim 11.

17. A method for developing inhibitors of aggrecanase comprising the use of aggrecanase protein set forth in SEQ ID NO. 7 or a fragment thereof.
 18. A method for developing inhibitors of aggrecanase comprising the use of the amino acid sequence set forth in SEQ ID NO:9 or a fragment thereof.
- 5
19. The method of claim 17 wherein said method comprises three dimensional structural analysis.
 20. The method of claim 18 wherein said method comprises computer aided drug design.
- 10
20. A composition for inhibiting the proteolytic activity of aggrecanase comprising a peptide molecule which binds to the aggrecanase inhibiting the proteolytic degradation of aggrecan.
 21. A method for inhibiting the cleavage of aggrecan in a mammal comprising administering to said mammal an effective amount of a compound that inhibits
- 15
- aggrecanase activity.
 22. An isolated nucleotide sequence comprising the DNA insert of ATCC PTA-2283.

23. A protein composition comprising the amino acid sequence set forth in SEQ ID NO:7 from nucleotide #140 to #1610 or a fragment thereof for use in the development of aggrecanase inhibitors.

- 5 24. A protein composition comprising the amino acid sequence set forth in SEQ ID NO:9 or a fragment thereof for use in the development of aggrecanase inhibitors.

SEQUENCE LISTING

<110> APPLICANT: Genetics Institute, Inc.

<120> TITLE: AGGRECANASE MOLECULES

<130> DOCKET/FILE REFERENCE: GI 5453

<150> PRIOR APPLICATION NUMBER: 60/243,916

<151> FILING DATE: 2000-10-27

<160> NUMBER OF SEQUENCES: 9

<170> SOFTWARE: FastSEQ for Windows Version 4.0

<210> SEQ ID NO:1

<211> LENGTH: 1506

<212> TYPE: DNA

<213> ORGANISM:Unknown

<220> FEATURE:

<223> OTHER INFORMATION: Unknown

<400> SEQ ID NO:1

```

aatcgtgggc cttccatatac ttttaatgct cagacaacat taaaaaactt ttgccagtgg 60
cagcattcga agaacagtcc aggtggaatc catcatgata ctgctgttct cttacaagg 120
tatatggatt tcttattcct cagatcttcc agtctcaggt caagttgatg ggatgtggtc 180
atgcattcct aatagtcagc tgctagtggg gggtttctca gcttaggcac tgttgacatt 240
ttgggctgga caattccttc tgtgcattgc aagatgttta acagcatccc tggcctctgc 300
tcactacgta ttagtagtac tccctcccct caggtgtgac aaccggcaat gtctccagac 360
attgccactg cccccgtgg ggagggcaga atcacctgg ttgagaacca cgaacctata 420
gataccaagt gaggaataaa tgtgttaggc agacctgaaa atacatgcct tatataacag 480
gcaagggcag ctgctgttca gatggacctg gcgttgccat gtagtatatg agttcagtgt 540
tgccagcttt tgtgattttg caaaagccag gaaaatgaat ttttaaaaaa tgtgaaacct 600
cccaagtttt ctatggtagc tactacctca aaatgttaaa aactgtagc aacaaagacg 660
cgtgctggct taattgtgcc tcccacctcc tgccatctca attgtggcca tgagcacgct 720
ctcttaaatg aaagtaaaac aaaggtatct tgcaccttct caggaatggt ctcaaagagt 780
ctttacgaaa ctggaatcat ttaccccagg gtattgggaa aatcctgggt ggagagattt 840
cttactgctc tgtgaccgcc ctgaccacag aagaaaactt cttaatgtag tagttataa 900
ttcttaaatc aacaattcca aaaatttttt gagttttgaa actctggtct tatgaatttg 960
ctttttttaa cccccacccc aaagtgtgta ctgttgtgaa tgcctagacy accctaaagt 1020
atttacttgt attgacattt actctttgat gtcacaaagg cctaagtctg ctttgcttc 1080
agacaggata tctgcagagc tcacgacaaa tgtgatacct taggcctggc tgaactggga 1140
accatttgtg atccctatag aagctgttct attagtgaag atagtggatt gactacagct 1200
tttacgatcg cccatgagct gggccatgtg taacatgcct catgatgaca acaacaaatg 1260
taaagaagaa ggagtttaaga gtcccagca tgtcatggct ccaacactga acttctacac 1320
caccccctgg atgtggtcaa agtgtagtcg aaaatatatc actgagtttt tagagtaaga 1380
cttgaacatt cttttagcac aaacttctag tgccctggcct acatgtagtg aactaattgt 1440
gggaaagaca atatgaagtc aacattcct tttgagttat tttgttgac attccttggg 1500
gaagac                                     1506

```

<210> SEQ ID NO:2

<211> LENGTH: 1028

<212> TYPE: DNA

<213> ORGANISM:unknown

<220> FEATURE:

<223> OTHER INFORMATION: Unknown

<400> SEQ ID NO:2

```

gaattcggcc aaagaggcct aatcatttat aggcgcagcg cccccagag agagccctca 60
acaggaaggc atgcatgtga cacctcagaa cacaaaaata ggcacagtaa agacaagaag 120
aaaaccagag caagaaaatg gggagaaagg attaacctgg ctggtgacgt agcagcatta 180

```

```

aacagcggct tagcaacaga ggcattttct gcttatggta ataagacgga caacacaaga 240
gaaaagagga cccacagaag gacaaaaagt tttttatcct atccacgggt tgtagaagtc 300
ttgggtgggtg cagacaacag aatggtttca taccatggag aaaaccttca acactatatt 360
ttaacttttaa tgtcaattgt agcctctatc tataaagacc caagtattgg aaattttaatt 420
aatattgtta ttgtgaactt aattgtgatt cataatgaac aggatgggccc ttccatatct 480
tttaatgctc agacaacatt aaaaaacttt tgccagtggc agcattcgaa gaacagttcca 540
ggtggaatcc atcatgatac tgctgtttctc ttaacaagac aggatatctg cagagctcac 600
gacaaatgtg ataccttagg cctgggtgaa ctggggaacca tttgtgatcc ctatagaagc 660
tgttctatta gtgaagatag tggattgagt acagctttta cgatcgccca tgagctgggc 720
catgtgttta acatgcctca tgatgacaac aacaaatgta aagaagaagg agttaagagt 780
ccccagcatg tcatggctcc aacactgaac ttctacacca acccctggat gtgggtcaaag 840
tgtagtcgaa aatatatcac tgagttttta gagtaagact tgaacattct tttagcacia 900
acttctagtg cctggcctac atgtagttaa ctaattgtgg gaaagacaat atgaagtcaa 960
acattccttt tgagttattt ttgttgacat tccttggaga aggcaaaaaa aaaaaaaaaa 1020
gcgggccgc                                     1028
    
```

```

<210> SEQ ID NO:3
<211> LENGTH: 1254
<212> TYPE: DNA
<213> ORGANISM: unknown
    
```

```

<220> FEATURE:
<223> OTHER INFORMATION: Unknown
    
```

```

<400> SEQ ID NO:3
aggggacaca tggcccagct catgggctgt cgtaaagctg tacccaatcc actatcttca 60
ctaatagaca gtttctatag gatcacaagt ggttcccagt tcagccaggc ctaaggatca 120
cattttgtcg tgagctctgc agatatacctg tctgaaagca aagcagactt agcctttgat 180
gacatcaaag agtaaagtgc aatacaagta aatactttag ggtcgtctag gcattcacia 240
cagtacacac ttgggggtgg ggggttaaaaa aagcaaatcc ataagaccag agtttcaaaa 300
ctcaaaaaat ttttgaatt gttgatttaa gaattataaa ctactacatt aagaagtttt 360
cttctgtggt cagggcggtc acagagcagt aagaaatccc tccaaccagg attttcccaa 420
taccctgggg taaatgattc cagtttcgta aagactcttt gagaacattc ctgagaaggt 480
gcaagatacc ttgttttac tttcatttaa gaggacgtgc tcatgaccac aattgagatg 540
gcaggagggtg ggaggcacia ttaagccagc acgcgtcttt gttgctacag tgttttaaca 600
ttttgagata gttagctacca tagaaaactt gggagggtttc acatttttaa aaaattcatt 660
tttctggctt ttgcaaaatc acaaaagctg gcaacactga actcatatac tcatggcaa 720
cgccaggctcc atctgaacag cagctgccct tgccgtttat ataaggcatg tattttcagg 780
tctgcctaac acatttattc ctcaactggg atctatagggt tctgtggtct caaccaaggt 840
gattctgccc tccccacggg gggcagtggt aatgtctgga gacattgctg gttgtcacac 900
ctgaggggag ggagtactac taatacgtag agagcagagg ccagggatgc tgtaaacat 960
cttgcaatgc acagaaggaa ttgtccagcc caaaatgtca acagtgccta agctgagaaa 1020
ccctccacta gcagctgact attaggaatg catggccaca tcccatcaac ttgacctgag 1080
actgaaagat ctgaggaata agaaatccat ataccctgtt aagagaacag cagtatcatg 1140
atggattcca cctggactgt tcttogaatg ctgccactgg caaaagtttt ttaatgtgt 1200
ctgagcatta aaagatatgg aaggcccacg attgaattct agacctgccc ccgc 1254
    
```

```

<210> SEQ ID NO:4
<211> LENGTH: 687
<212> TYPE: DNA
<213> ORGANISM: unknown
    
```

```

<220> FEATURE:
<223> OTHER INFORMATION: Unknown
    
```

```

<400> SEQ ID NO:4
ggccaaagag gcctaccatg gagaaaacct tcaacactat attttaactt taatgtcaat 60
tgtagcctct atctataaag acccaagtat tggaaattta attaataattg ttattgtgaa 120
cttaattgtg attcataatg aacaggatgg gccttccata tcttttaatg ctcagaciac 180
attaaaaaac ttttgccagt ggcagcattc gaagaacagt ccaggtggaa tccatcatga 240
tactctgtgt ctcttaacaa gacaggatat ctgcagagct cagcaciaat gtgatacctt 300
aggcctggct gaactgggaa ccatttgtga tccctataga agctgttcta ttagtgaaga 360
tagtggattg agtacagctt ttacgatcgc ccatgagctg ggccatgtgt ttaacatgcc 420
tcatagctag aacaacaaat gtaaagaaga aggagtcaag agtccccagc atgtcatggc 480
tccaactatg aacttctaca ccaaccctg gatgtgggtc aagtgtagtc gaaaatatat 540
cgctgagttt tttagagtaa acttgaacat tcttttagca caaacttcta gtgctggccc 600
    
```

tacatgtagt gaactaattg tgggaaagac aatatgaagt caaacattcc ttttgagtta 660
 tttttgttga cattccttgg agaagggc 687

<210> SEQ ID NO:5
 <211> LENGTH: 466
 <212> TYPE: DNA
 <213> ORGANISM:unknown

<220> FEATURE:
 <223> OTHER INFORMATION: Unknown

<400> SEQ ID NO:5
 ggccaaagag gcctacactg ctaacgctcc tgggtgcggga cctggccgag atggggagcc 60
 cagacgccgc ggcggccgtg cgcaaggaca ggctgcaccc gaggcaagtg aaattattag 120
 agaccctgag cgaatacga atcgtgtctc ccatccgagt gaacgctctc ggagaaccct 180
 ttcccacgaa cgtccacttc aaaagaacgc gacggagcat taactctgcc actgaccct 240
 ggctgcctt cgctcctcc tcttctctc ctacctctc ccaggcgc ataccgctct 300
 ctgccttcgg ccagcagttt ctatttaatc tcaccgcaa tgccggattt atcgctccac 360
 tgttcactgt caccctctc gggacgccc ggggtgaatca gaccaagttt tattccgaag 420
 aggaagcgga actcaagcac tgtttctaca aaggctatgt caatac 466

<210> SEQ ID NO:6
 <211> LENGTH: 5605
 <212> TYPE: DNA
 <213> ORGANISM:unknown

<220> FEATURE:
 <223> OTHER INFORMATION: Unknown

<400> SEQ ID NO:6
 ggccaaagag gcctacactg ctaacgctcc tgggtgcggga cctggccgag atggggagcc 60
 cagacgccgc ggcggccgtg cgcaaggaca ggctgcaccc gaggcaagtg aaattattag 120
 agaccctgag cgaatacga atcgtgtctc ccatccgagt gaacgctctc ggagaaccct 180
 ttcccacgaa cgtccacttc aaaagaacgc gacggagcat taactctgcc actgaccct 240
 ggctgcctt cgctcctcc tcttctctc ctacctctc ccaggcgc ataccgctct 300
 ctgccttcgg ccagcagttt ctatttaatc tcaccgcaa tgccggattt atcgctccac 360
 tgttcactgt caccctctc gggacgccc ggggtgaatca gaccaagttt tattccgaag 420
 aggaagcgga actcaagcac tgtttctaca aaggctatgt caataccaac tccgagcaca 480
 cggccgtcat cagcctctgc tcaggaatgc tgggcacatt ccggtctcat gatggggatt 540
 attttattga accactacag tctatggatg aacaagaaga tgaagaggaa caaaacaaac 600
 cccacatcat ttataggcgc agcgcacccc agagagagcc ctcaacagga aggcattgcat 660
 gtgacacctc agaacacaaa aataggcaca gtaaagacaa gaagaaaacc agagcaagaa 720
 aatggggaga aaggattaac ctggctgggtg acgtagcagc attaaacagc ggcttagcaa 780
 cagagcattt ttctgcttat ggtaataaga cggacaacac aagagaaaag aggccccaca 840
 gaaggacaaa acgtttttta tcctatccac ggtttgtaga agtcttgggtg gtggcagaca 900
 acagaatggt ttcataccat ggagaaaacc ttcaacacta tattttaact ttaatgtcaa 960
 ttgtagcctc tatctataaa gaccaagta ttggaaatth aattaatatt gttattgtga 1020
 acttaattgt gattcataat gaacaggatg ggccttccat atcttttaac gctcagacaa 1080
 cattaataaa cttttgccag tggcagcatt cgaagaacag tccagggtgga atccatcatg 1140
 atactgctgt tctcttaaca agacaggata tctgcagagc tcacgacaaa tgtgatacct 1200
 taggctggc tgaactggga accatttctg atccctatag aagctgttct attagtgaag 1260
 atagtggatt gactacagct tttacgatcg cccatgagct gggccatgtg ttaacatgc 1320
 ctcatgatga caacaacaaa tgtaagaag aaggagttaa gagtcccag catgtcatgg 1380
 ctccaacact gaacttctac accaaccctt ggatgtgggtc aaagtgtagt cgaaaatata 1440
 tcaactgagt ttttagacact ggttatggcg agtgtttgct taacgaacct gaatccagac 1500
 cctacccttt gctgtccaa ctgccaggca tcttttaciaa cgtgaataaa caatgtgaat 1560
 tgatthttgg accaggttct cagggtgtgcc catatatgat gcagtgcaga cggctctggt 1620
 gcaataacgt caatggagta cacaaaggct gccggactca gcacacaccc tgggcccgatg 1680
 ggacggagtg cgagcctgga aagcactgca agtatggatt ttgtgttccc aaagaaatgg 1740
 atgtccccgt gacagatgga tctgggggaa gttggagttc ctttggaaacc tgctccagaa 1800
 catgtggagg gggcatcaaa acagccattc gagagtgcaa cagaccagaa ccaaaaaatg 1860
 gtggaaaata ctgtgttaga cgtagaatga aatttaagtc ctgcaacacg gagccatgtc 1920
 tcaagcagaa gcgagacttc cgagatgaac agtgtgtctc ctttgacggg aagcatttta 1980
 acatcaacgg tctgcttccc gggtccttaa atacagtga atctgtgga 2040
 aggaccggtg caagttgttc tgcagagtgg cagggaacac agcctactat cagcttcgag 2100

acagagtgat	agatggaact	ccttgtggcc	aggacacaaa	tgatatctgt	gtccagggcc	2160
tttgccggca	agctggatgc	gatcatgttt	taaactcaaa	agcccggaga	gataaatgtg	2220
gggtttgtgg	tggcgataat	tcttcatgca	aaacagtggc	aggaacattt	aatacagtac	2280
attatggtta	caatactgtg	gtccgaattc	caagtgtgtc	taccaatatt	gatgtgaggc	2340
agcacagttt	ctcaggggaa	acagacgatg	acaactactt	agctttatca	agcagtaaag	2400
gtgaattcct	gctaaatgga	aactttgttg	tcacaatggc	caaaagggaa	attcgcattg	2460
ggaatgctgt	ggtagagtac	agtgggtccg	agactgccgt	agaaagaatt	aactcaacag	2520
atcgcattga	gcaagaactt	ttgcttcagg	ttttgtcgg	gggaaagtgt	tacaaccccc	2580
atgtacgcta	ttctttcaat	attccaattg	aagataaacc	tcagcagttt	tactggaaca	2640
gtcatgggcc	atggcaagca	tgcagtaaac	cctgcccaagg	ggaacgggaaa	cgaaaacttg	2700
tttgcaccag	ggaatctgat	cagcttactg	tttctgatca	aagatgcat	cggtgcccc	2760
agcctggaca	cattactgaa	ccctgtggta	cagactgtga	cctgaggtgg	catgttgcca	2820
gcaggagtga	atgtagtgcc	cagtgtggct	tgggtaccg	cacattggac	atctactgtg	2880
ccaaatatag	caggctggat	gggaagactg	agaaggttga	tgatggtttt	tgcagcagcc	2940
atcccaaacc	aagcaaccgt	gaaaaatgct	caggggaatg	taacacgggt	ggctggcgct	3000
attctgcctg	gactgaatgt	tcaaaaagct	gtgacggtgg	gacccagagg	agaagggcta	3060
ttctgtcctc	taccgaaat	gatgtactgg	atgcagcaa	atgcacacat	caagagaaag	3120
ttaccattca	gaggtgcagt	gagttccctt	gtccacagtg	gaaatctgga	gactggctcag	3180
agtgtctgg	cacctgtgga	aaagggcata	agcaccgcca	ggtctgggtg	cagtttggtg	3240
aagatcgatt	aaatgataga	atgtgtgacc	ctgagaccaa	gccaacatct	atgcagactt	3300
gtcagcagcc	ggaatgtgca	tcctggcagg	cggtccctg	gggacagtgc	agtgtcactt	3360
gtggacaggg	ataccagcta	agagcagtga	aatgcatcat	tgggacttat	atgtcagttg	3420
tagatgacaa	tgactgtaat	gcagcaacta	gaccaactga	taccaggac	tgtgaattac	3480
catcatgtca	tcctcccca	gctgccccg	aaacgaggag	aagcacatac	agtgcaccaa	3540
gaaccagtg	gcgatttggg	tcttggacc	catgtcagc	cacttgtggg	aaaggtacc	3600
ggatgagata	cgtcagctgc	cgagatgaga	atggctctgt	ggctgacgag	agtgcctgtg	3660
ctaccctgcc	tagaccagtg	gcaaaggaag	aatgttctgt	gacaccctgt	gggcaatgga	3720
aggccttgg	ctggagctct	tgctctgtga	cctgtgggca	aggtagggca	accgggcaag	3780
tgatgtgtgt	caactacagt	gaccacgtga	tcgatcggag	tgagtgtgac	caggattata	3840
tcccagaaac	tgaccaggac	tgttccatgt	caccatgccc	tcaaaggacc	ccagacagtg	3900
gcttagctca	gcaccccttc	caaaatgagg	actatcgcc	ccggagcgcc	agccccagcc	3960
gcacccatgt	gctcggtgga	aaccagtgga	gaactggccc	ctggggagca	tgttccagta	4020
cctgtgctgg	cgatcccag	cggtgtgtg	ttgtatgtca	ggatgaaaat	ggatacaccg	4080
caaacgactg	tgtggagaga	ataaaacctg	atgagcaaag	agcctgtgaa	tccggccctt	4140
gtcctcagtg	ggcttatggc	aactggggag	agtgcactaa	gctgtgtggt	ggaggcataa	4200
gaacaagact	tgtgtgtgt	cagcgggtcca	acgggtgaacg	gtttccagat	ttgagctgtg	4260
aaattcctga	taaactctcc	gatcgtgagc	agtgtaacac	acatgcttgt	ccacacgacg	4320
ctgcatggag	tactggccct	tggagctcgt	gttctgtctc	ttgtgggtcga	gggcataaac	4380
aacgaaatgt	ttactgcatg	gcaaaagatg	gaagccattt	agaaagtgat	tactgtaagc	4440
acctgtcctaa	gacacatggg	cacagaaagt	gccgaggagg	aagatgcccc	aaatggaaag	4500
ctggcgcttg	gagtcagtg	tctgtgtcct	gtggccgagg	cgtacagcag	aggcatgtgg	4560
gctgtcagat	cggaacacac	aaaatagcca	gagagaccga	gtgcaaccca	tacaccagac	4620
cgtagtcgga	acgcgactgc	caagggccac	ggtgtcccct	ctacacttgg	agggcagagg	4680
aatggcaaga	atgcaccaag	acctgcggcg	aaggctccag	gtaccgcaag	gtgggtgtgtg	4740
tggatgacaa	caaaaacgag	gtgcatgggg	cacgctgtga	cgtgagcaag	cgccgggtgg	4800
accgtgaaag	ctgtagtgtg	caaccctgcg	agtatgtctg	gatcacagga	gaatggctcag	4860
aggtaccgct	ctgggaactg	taaccatcgt	cagctcagcc	atggcctgag	agtggcagag	4920
ggatgagtgg	agggatgagt	gcaggaatgt	gggagacttg	aggctaccgg	cccgatttgc	4980
cactgtgaac	tgtgtgtttt	ctgacaagtc	ctcagcttcc	ccaagctaga	attccttgta	5040
tgcaaagcgg	gagagatgta	agagatggtc	tctaagtccc	ttcaggtcta	cattctgtga	5100
ttcaccttga	tgtcctattg	gcataaagaa	gaaattatta	caggggctgc	aaactcatag	5160
gatgctgtga	ggtgcctgaa	gacagttaag	tataagaaaa	tattgtagtg	ccagggatac	5220
acaaggaga	gatggcaact	gtgacaaact	agcacatgct	gtgtgaaggg	agcagaatct	5280
ctttcactcc	agctgtggcc	atgcagaaat	gtggctctagc	gttaccagac	ctgatttttc	5340
aagagaggct	aaaaatctgg	actagtatgt	gagatttcc	aacttgaaaa	tgggggctga	5400
aatttttgg	tttaaaacat	tgtaaggggc	aaacaaacc	ctttcatgaa	ccagatgtgt	5460
tgtgcctgtt	taacaaacag	cttcagagga	agaaaataat	tttctataat	atccgaagta	5520
tctcaagtac	cattttttca	tatatcttcc	tgtgcacaat	gcttatctag	acccttttta	5580
atggttaata	accagtagta	atcat				5605

<210> SEQ ID NO:7
 <211> LENGTH: 1602
 <212> TYPE: PRT
 <213> ORGANISM:unknown

<220> FEATURE:
 <223> OTHER INFORMATION: Unknown

<400> SEQ ID NO:7

Met	Gly	Ser	Pro	Asp	Ala	Ala	Ala	Ala	Val	Arg	Lys	Asp	Arg	Leu	His
1				5					10					15	
Pro	Arg	Gln	Val	Lys	Leu	Leu	Glu	Thr	Leu	Ser	Glu	Tyr	Glu	Ile	Val
			20					25					30		
Ser	Pro	Ile	Arg	Val	Asn	Ala	Leu	Gly	Glu	Pro	Phe	Pro	Thr	Asn	Val
		35					40					45			
His	Phe	Lys	Arg	Thr	Arg	Arg	Ser	Ile	Asn	Ser	Ala	Thr	Asp	Pro	Trp
	50					55					60				
Pro	Ala	Phe	Ala	Ser	Ser	Ser	Ser	Ser	Ser	Thr	Ser	Ser	Gln	Ala	His
65					70					75					80
Tyr	Arg	Leu	Ser	Ala	Phe	Gly	Gln	Gln	Phe	Leu	Phe	Asn	Leu	Thr	Ala
				85					90					95	
Asn	Ala	Gly	Phe	Ile	Ala	Pro	Leu	Phe	Thr	Val	Thr	Leu	Leu	Gly	Thr
			100					105					110		
Pro	Gly	Val	Asn	Gln	Thr	Lys	Phe	Tyr	Ser	Glu	Glu	Glu	Ala	Glu	Leu
		115					120					125			
Lys	His	Cys	Phe	Tyr	Lys	Gly	Tyr	Val	Asn	Thr	Asn	Ser	Glu	His	Thr
	130					135					140				
Ala	Val	Ile	Ser	Leu	Cys	Ser	Gly	Met	Leu	Gly	Thr	Phe	Arg	Ser	His
145					150					155					160
Asp	Gly	Asp	Tyr	Phe	Ile	Glu	Pro	Leu	Gln	Ser	Met	Asp	Glu	Gln	Glu
				165					170					175	
Asp	Glu	Glu	Glu	Gln	Asn	Lys	Pro	His	Ile	Ile	Tyr	Arg	Arg	Ser	Ala
			180					185					190		
Pro	Gln	Arg	Glu	Pro	Ser	Thr	Gly	Arg	His	Ala	Cys	Asp	Thr	Ser	Glu
		195					200					205			
His	Lys	Asn	Arg	His	Ser	Lys	Asp	Lys	Lys	Lys	Thr	Arg	Ala	Arg	Lys
	210					215					220				
Trp	Gly	Glu	Arg	Ile	Asn	Leu	Ala	Gly	Asp	Val	Ala	Ala	Leu	Asn	Ser
225					230					235					240
Gly	Leu	Ala	Thr	Glu	Ala	Phe	Ser	Ala	Tyr	Gly	Asn	Lys	Thr	Asp	Asn
				245					250					255	
Thr	Arg	Glu	Lys	Arg	Thr	Arg	Thr	Lys	Arg	Phe	Leu	Ser	Tyr	Pro	Arg
			260					265					270		
Phe	Val	Glu	Val	Leu	Val	Val	Ala	Asp	Asn	Arg	Met	Val	Ser	Tyr	His
		275					280					285			
Gly	Glu	Asn	Leu	Gln	His	Tyr	Ile	Leu	Thr	Leu	Met	Ser	Ile	Val	Ala
	290					295					300				
Ser	Ile	Tyr	Lys	Asp	Pro	Ser	Ile	Gly	Asn	Leu	Ile	Asn	Ile	Val	Ile
305					310					315					320
Val	Asn	Leu	Ile	Val	Ile	His	Asn	Glu	Gln	Asp	Gly	Pro	Ser	Ile	Ser
				325					330					335	
Phe	Asn	Ala	Gln	Thr	Thr	Leu	Lys	Asn	Phe	Cys	Gln	Trp	Gln	His	Ser
			340					345					350		
Lys	Asn	Ser	Pro	Gly	Gly	Ile	His	His	Asp	Thr	Ala	Val	Leu	Leu	Thr
		355					360					365			
Arg	Gln	Asp	Ile	Cys	Arg	Ala	His	Asp	Lys	Cys	Asp	Thr	Leu	Gly	Leu
	370					375					380				
Ala	Glu	Leu	Gly	Thr	Ile	Cys	Asp	Pro	Tyr	Arg	Ser	Cys	Ser	Ile	Ser
385					390					395					400
Glu	Asp	Ser	Gly	Leu	Ser	Thr	Ala	Phe	Thr	Ile	Ala	His	Glu	Leu	Gly
				405					410					415	
His	Val	Phe	Asn	Met	Pro	His	Asp	Asp	Asn	Asn	Lys	Cys	Lys	Glu	Glu
			420					425					430		
Gly	Val	Lys	Ser	Pro	Gln	His	Val	Met	Ala	Pro	Thr	Leu	Asn	Phe	Tyr
		435					440						445		

Thr	Asn	Pro	Trp	Met	Trp	Ser	Lys	Cys	Ser	Arg	Lys	Tyr	Ile	Thr	Glu
	450					455					460				
Phe	Leu	Asp	Thr	Gly	Tyr	Gly	Glu	Cys	Leu	Leu	Asn	Glu	Pro	Glu	Ser
465				470							475				480
Arg	Pro	Tyr	Pro	Leu	Pro	Val	Gln	Leu	Pro	Gly	Ile	Leu	Tyr	Asn	Val
				485					490					495	
Asn	Lys	Gln	Cys	Glu	Leu	Ile	Phe	Gly	Pro	Gly	Ser	Gln	Val	Cys	Pro
			500					505					510		
Tyr	Met	Met	Gln	Cys	Arg	Arg	Leu	Trp	Cys	Asn	Asn	Val	Asn	Gly	Val
		515					520					525			
His	Lys	Gly	Cys	Arg	Thr	Gln	His	Thr	Pro	Trp	Ala	Asp	Gly	Thr	Glu
	530					535					540				
Cys	Glu	Pro	Gly	Lys	His	Cys	Lys	Tyr	Gly	Phe	Cys	Val	Pro	Lys	Glu
545				550						555					560
Met	Asp	Val	Pro	Val	Thr	Asp	Gly	Ser	Trp	Gly	Ser	Trp	Ser	Pro	Phe
				565					570					575	
Gly	Thr	Cys	Ser	Arg	Thr	Cys	Gly	Gly	Gly	Ile	Lys	Thr	Ala	Ile	Arg
				580					585				590		
Glu	Cys	Asn	Arg	Pro	Glu	Pro	Lys	Asn	Gly	Gly	Lys	Tyr	Cys	Val	Gly
		595					600					605			
Arg	Arg	Met	Lys	Phe	Lys	Ser	Cys	Asn	Thr	Glu	Pro	Cys	Leu	Lys	Gln
	610					615						620			
Lys	Arg	Asp	Phe	Arg	Asp	Glu	Gln	Cys	Ala	His	Phe	Asp	Gly	Lys	His
625					630					635					640
Phe	Asn	Ile	Asn	Gly	Leu	Leu	Pro	Asn	Val	Arg	Trp	Val	Pro	Lys	Tyr
				645					650					655	
Ser	Gly	Ile	Leu	Met	Lys	Asp	Arg	Cys	Lys	Leu	Phe	Cys	Arg	Val	Ala
			660					665					670		
Gly	Asn	Thr	Ala	Tyr	Tyr	Gln	Leu	Arg	Asp	Arg	Val	Ile	Asp	Gly	Thr
		675				680						685			
Pro	Cys	Gly	Gln	Asp	Thr	Asn	Asp	Ile	Cys	Val	Gln	Gly	Leu	Cys	Arg
	690					695					700				
Gln	Ala	Gly	Cys	Asp	His	Val	Leu	Asn	Ser	Lys	Ala	Arg	Arg	Asp	Lys
705					710					715					720
Cys	Gly	Val	Cys	Gly	Gly	Asp	Asn	Ser	Ser	Cys	Lys	Thr	Val	Ala	Gly
				725					730					735	
Thr	Phe	Asn	Thr	Val	His	Tyr	Gly	Tyr	Asn	Thr	Val	Val	Arg	Ile	Pro
			740					745					750		
Ala	Gly	Ala	Thr	Asn	Ile	Asp	Val	Arg	Gln	His	Ser	Phe	Ser	Gly	Glu
		755					760					765			
Thr	Asp	Asp	Asp	Asn	Tyr	Leu	Ala	Leu	Ser	Ser	Ser	Ser	Lys	Gly	Glu
	770					775						780			Phe
Leu	Leu	Asn	Gly	Asn	Phe	Val	Val	Thr	Met	Ala	Lys	Arg	Glu	Ile	Arg
785					790					795					800
Ile	Gly	Asn	Ala	Val	Val	Glu	Tyr	Ser	Gly	Ser	Glu	Thr	Ala	Val	Glu
				805					810					815	
Arg	Ile	Asn	Ser	Thr	Asp	Arg	Ile	Glu	Gln	Glu	Leu	Leu	Leu	Gln	Val
			820					825					830		
Leu	Ser	Val	Gly	Lys	Leu	Tyr	Asn	Pro	Asp	Val	Arg	Tyr	Ser	Phe	Asn
		835					840					845			
Ile	Pro	Ile	Glu	Asp	Lys	Pro	Gln	Gln	Phe	Tyr	Trp	Asn	Ser	His	Gly
	850					855					860				
Pro	Trp	Gln	Ala	Cys	Ser	Lys	Pro	Cys	Gln	Gly	Glu	Arg	Lys	Arg	Lys
865					870					875					880
Leu	Val	Cys	Thr	Arg	Glu	Ser	Asp	Gln	Leu	Thr	Val	Ser	Asp	Gln	Arg
				885					890					895	
Cys	Asp	Arg	Leu	Pro	Gln	Pro	Gly	His	Ile	Thr	Glu	Pro	Cys	Gly	Thr
			900					905					910		
Asp	Cys	Asp	Leu	Arg	Trp	His	Val	Ala	Ser	Arg	Ser	Glu	Cys	Ser	Ala
		915					920					925			
Gln	Cys	Gly	Leu	Gly	Tyr	Arg	Thr	Leu	Asp	Ile	Tyr	Cys	Ala	Lys	Tyr
	930					935					940				
Ser	Arg	Leu	Asp	Gly	Lys	Thr	Glu	Lys	Val	Asp	Asp	Gly	Phe	Cys	Ser
945					950					955					960
Ser	His	Pro	Lys	Pro	Ser	Asn	Arg	Glu	Lys	Cys	Ser	Gly	Glu	Cys	Asn
				965					970					975	

Thr Gly Gly Trp Arg Tyr Ser Ala Trp Thr Glu Cys Ser Lys Ser Cys
 980 985 990
 Asp Gly Gly Thr Gln Arg Arg Arg Ala Ile Cys Val Asn Thr Arg Asn
 995 1000 1005
 Asp Val Leu Asp Asp Ser Lys Cys Thr His Gln Glu Lys Val Thr Ile
 1010 1015 1020
 Gln Arg Cys Ser Glu Phe Pro Cys Pro Gln Trp Lys Ser Gly Asp Trp
 1025 1030 1035 1040
 Ser Glu Cys Leu Val Thr Cys Gly Lys Gly His Lys His Arg Gln Val
 1045 1050 1055
 Trp Cys Gln Phe Gly Glu Asp Arg Leu Asn Asp Arg Met Cys Asp Pro
 1060 1065 1070
 Glu Thr Lys Pro Thr Ser Met Gln Thr Cys Gln Gln Pro Glu Cys Ala
 1075 1080 1085
 Ser Trp Gln Ala Gly Pro Trp Gly Gln Cys Ser Val Thr Cys Gly Gln
 1090 1095 1100
 Gly Tyr Gln Leu Arg Ala Val Lys Cys Ile Ile Gly Thr Tyr Met Ser
 1105 1110 1115 1120
 Val Val Asp Asp Asn Asp Cys Asn Ala Ala Trp Thr Asp Thr Gln Asp
 1125 1130 1135
 Cys Glu Leu Pro Ser Cys His Pro Pro Ala Ala Pro Glu Thr Arg
 1140 1145 1150
 Arg Ser Thr Tyr Ser Ala Pro Arg Thr Gln Trp Arg Phe Gly Ser Trp
 1155 1160 1165
 Thr Pro Cys Ser Ala Thr Cys Gly Lys Gly Thr Arg Met Arg Tyr Val
 1170 1175 1180
 Ser Cys Arg Asp Glu Asn Gly Ser Val Ala Asp Glu Ser Ala Cys Ala
 1185 1190 1195 1200
 Thr Leu Pro Arg Pro Val Ala Lys Glu Glu Cys Ser Val Thr Pro Cys
 1205 1210 1215
 Gly Gln Trp Lys Ala Leu Asp Trp Ser Ser Cys Ser Val Thr Cys Gly
 1220 1225 1230
 Gln Gly Arg Ala Thr Arg Gln Val Met Cys Val Asn Tyr Ser Asp His
 1235 1240 1245
 Val Ile Asp Arg Ser Glu Cys Asp Gln Asp Tyr Ile Pro Glu Thr Asp
 1250 1255 1260
 Gln Asp Cys Ser Met Ser Pro Cys Pro Gln Arg Thr Pro Asp Ser Gly
 1265 1270 1275 1280
 Leu Ala Gln His Pro Phe Gln Asn Glu Asp Tyr Arg Pro Arg Ser Asp
 1285 1290 1295
 Ser Arg Thr His Val Leu Gly Gly Asn Gln Trp Arg Thr Gly Pro Trp
 1300 1305 1310
 Gly Ala Cys Ser Ser Thr Cys Ala Gly Gly Ser Gln Arg Arg Val Val
 1315 1320 1325
 Val Cys Gln Asp Glu Asn Gly Tyr Thr Ala Asn Asp Cys Val Glu Arg
 1330 1335 1340
 Ile Lys Pro Asp Glu Gln Arg Ala Cys Glu Ser Gly Pro Cys Pro Gln
 1345 1350 1355 1360
 Trp Ala Tyr Gly Asn Trp Gly Glu Cys Thr Lys Leu Cys Gly Gly Gly
 1365 1370 1375
 Ile Arg Thr Arg Leu Val Val Cys Gln Arg Ser Asn Gly Glu Arg Phe
 1380 1385 1390
 Pro Asp Leu Ser Cys Glu Ile Leu Asp Lys Pro Pro Asp Arg Glu Gln
 1395 1400 1405
 Cys Asn Thr His Ala Cys Pro His Asp Ala Ala Trp Ser Thr Gly Pro
 1410 1415 1420
 Trp Ser Ser Cys Ser Val Ser Cys Gly Arg Gly His Lys Gln Arg Asn
 1425 1430 1435 1440
 Val Tyr Cys Met Ala Lys Asp Gly Ser His Leu Glu Ser Asp Tyr Cys
 1445 1450 1455
 Lys His Leu Ala Lys Pro His Gly His Arg Lys Cys Arg Gly Gly Arg
 1460 1465 1470
 Cys Pro Lys Trp Lys Ala Gly Ala Trp Ser Gln Cys Ser Val Ser Cys
 1475 1480 1485
 Gly Arg Gly Val Gln Gln Arg His Val Gly Cys Gln Ile Gly Thr His
 1490 1495 1500

Lys Ile Ala Arg Glu Thr Glu Cys Asn Pro Tyr Trp Glu Ser Glu Arg
 1505 1510 1515 1520
 Asp Cys Gln Gly Pro Arg Cys Pro Leu Tyr Thr Trp Arg Ala Glu Glu
 1525 1530 1535
 Trp Gln Glu Cys Thr Lys Thr Cys Gly Glu Gly Ser Arg Tyr Arg Lys
 1540 1545 1550
 Val Val Cys Val Asp Asp Asn Lys Asn Glu Val His Gly Ala Arg Cys
 1555 1560 1565
 Asp Val Ser Lys Arg Pro Val Asp Arg Glu Ser Cys Ser Leu Gln Pro
 1570 1575 1580
 Cys Glu Tyr Val Trp Ile Thr Gly Glu Trp Ser Glu Val Pro Ser Trp
 1585 1590 1595 1600
 Glu Leu

<210> SEQ ID NO:8
 <211> LENGTH: 5053
 <212> TYPE: DNA
 <213> ORGANISM: unknown

<220> FEATURE:
 <223> OTHER INFORMATION: Unknown

<400> SEQ ID NO:8
 cgggccgcca ccatgcagtt tgtatcctgg gccacactgc taacgctcct ggtgcgggac 60
 ctggccgaga tggggagccc agacgccgag gggcccgctg gcaaggacag gctgcaccgg 120
 aggcaagtga aattattgga gaccctgagc gaatacgaaa tcgtgtcttc catccgagtg 180
 aacgctctcg gagaaccctt tcccacgaac gtccacttca aaagaacgag acggagcatt 240
 aactctgcca ctgacccttg gcctgccttc gcctcctcct cttcctcctc tacctccacc 300
 caggcgcatt accgcctctc tgccttgggc cagcagtttc tatttaatct caccgccaat 360
 gccggattta tcgctccact gttcactgtc accctcctcg ggacgcccgg ggtgaatcag 420
 accaagtttt attccgaaga ggaagcggaa ctcaagcact gtttctacaa aggctatgtc 480
 aataccaact ccgagcacac ggccgtcatc agcctctgct caggaatgct gggcacattc 540
 cggctctcatg atggggatta ttttattgaa ccactacagt ctatggatga acaagaagat 600
 gaagaggaac aaaacaaacc ccacatcatt tataggcgca gcgccccca gagagagccc 660
 tcaacaggaa ggcatgcatg tgacacctca gaacacaaaa ataggcacag taaagacaag 720
 aagaaaacca cagcaagaaa atggggagaa aggcattacc tggctgggta cgtagcagca 780
 ttaaacagcg gcttagcaac agaggcattt tctgcttatg gtaataagac ggacaacaca 840
 agagaaaaga ggaccacacag aaggacaaaa cgttttttat cctatccacy gttttagtagaa 900
 gtcttgggtg tggcagacaa cagaatggtt tcataccatg gagaaaacct tcaacactat 960
 attttaactt taatgtcaat tgtagcctct atctataaag acccaagtat tggaaattta 1020
 attaataattg ttattgtgaa ctttaattgtg attcataatg aacaggatgg gccttccata 1080
 tcttttaattg ctcagacaac attaaaaaac ttttgccagt ggcagcattc gaagaacagt 1140
 ccagggtgaa tccatcatga tactgctggt cctttaacaa gacaggatat ctgcagagct 1200
 cagcaaaaat gtgatacctt aggcctggct gaactgggaa ccatttggta tcctataga 1260
 agctgttcta ttagtgaaga tagtggattg agtacagctt ttacgatcgc ccatgagctg 1320
 ggccatgtgt ttaacatgcc tcatgatgac aacaacaaat gtaaaagaag aggagttaag 1380
 agtccccagc atgtcatggc tccaacactg aacttctaca ccaaccctcg gatgtggtca 1440
 aagtgtagtc gaaaatatat cactgagttt ttagacactg gttatggcga gtgtttgctt 1500
 aacgaacctg aatccagacc ctaccctttg cctgtccaac tgccaggcat cctttacaac 1560
 gtgaataaac aatgtgaatt gatttttggg ccaggttctc aggtgtgccc atatatgatg 1620
 cagtgcagac ggctctggtg caataacgtc aatggagtac acaaaggctg ccggactcag 1680
 cacacaccct gggccgatgg gacggagtgc gagcctggaa agcactgcaa gtatggattt 1740
 tgtgttccca aagaaatgga tgtccccgtg acagatggat cctggggaag ttggagtcct 1800
 tttggaacct gctccagaac atgtggaggg ggcatacaaaa cagccattcg agagtgcaac 1860
 agaccagaac caaaaaatgg tggaaaatac tgtgtaggac gtagaatgaa atttaagtcc 1920
 tgcaacacgg agccatgtct caagcagaag cgagacttcc gagatgaaca gtgtgctcac 1980
 tttgacggga agcattttta catcaacggg ctgcttccca atgtgcgctg ggtccctaaa 2040
 tacagtggaa ttctgatgaa ggaccgggtc aagttgttct gcagagtggc agggaacaca 2100
 gcctactatc agcttcgaga cagagtgata gatggaaact cttgtgggtca ggacacaaa 2160
 gatattctgtg tccagggcct ttgccggcaa gctggatgag atcatgtttt aaactcaaaa 2220
 gcccgagag ataaatgtgg ggtttgtggt ggcgataaatt cttcatgcaa aacagtggca 2280
 ggaacattta atacagtaca ttatggttac aatactgtgg tccgaattcc agctggtgct 2340
 accaatattg atgtgcggca gcacagtttc tcaggggaaa cagacgatga caactactta 2400
 gctttatcaa gcagtaaagg tgaattcttg ctaaatggaa actttgttgt cacaatggcc 2460

```

aaaagggaaa ttcgcattgg gaatgctgtg gtagagtaca gtgggtccga gactgccgta 2520
gaaagaatta actcaacaga tcgcattgag caagaacttt tgcttcaggt tttgtcgggt 2580
ggaaagtgtg acaaccccga tgtacgctat tctttcaata ttccaattga agataaacct 2640
cagcagtttt actggaacag tcatgggcca tggcaagcat gcagtaaacc ctgccaaggg 2700
gaacggaaac gaaaacttgt ttgcaccagg gaatctgatc agcttactgt ttctgatcaa 2760
agatgcgatc ggctgccccca gcctggacac attactgaac cctgtgggtac agactgtgac 2820
ctgaggtggc atgttgccag caggagtgaa tgtagtgcc agtgtggctt gggttaccgc 2880
acattggaca tctactgtgc caaataatagc aggctggatg ggaagactga gaaggttgat 2940
gatgggtttt gcagcagcca tcccaaacca agcaaccgtg aaaaatgctc aggggaaatgt 3000
aacacgggtg gctggcgcta ttctgcctgg actgaaatgtt caaaaagctg tgacgggtggg 3060
accagagga gaagggctat ttgtgtcaat acccgaaatg atgtactgga tgacagcaaa 3120
tgcacacatc aagagaaagt taccattcag aggtgcagtg agttcccttg tccacagtgg 3180
aaatctggag actggtcaga gtgcttggtc acctgtggaa aagggcataa gcaccgccag 3240
gtctgggtgc agtttggtga agatcgatta aatgatagaa tgtgtgacct tgagaccaag 3300
ccaactcta tgcagacttg tcagcagccg gaatgtgcat cctggcagggc gggttccctgg 3360
ggacagtgca gtgtcacttg tggacagga taccagctaa gagcagtgaa atgcatcatt 3420
gggacttata tgtcagtggt agatgacaat gactgtaatg cagcaactag accaactgat 3480
accaggact gtgaattacc atcatgtcat cctccccag ctgccccgga aacgaggaga 3540
agcacatata gtgcaccaag aaccagtggt cgatttgggt cttggacccc atgtcagcc 3600
acttgtggga aaggtaccgg gatgagatac gtcagctgcc gagatgagaa tggctctgtg 3660
gctgacgaga gtgcctgtgc taccctgcct agaccagtggt caaaggaaga atgttctgtg 3720
acaccctgtg gccaatggaa ggccttggac tggagctctt gctctgtgac ctgtgggcaa 3780
ggtagggcaa cccggcaagt gatgtgtgc aactacagtg accacgtgat cgatcggagt 3840
gagtgtgacc aggattatat ccagaaact gaccaggact gttccatgtc accatgccct 3900
caaaggaccc cagacagtgg cttagctcag cccccctcc aaaatgagga ctatcgtccc 3960
cggagcggca ccccagccg caccatgtg ctcggtggaa accagtggag aactggcccc 4020
tgaggagcat gttccagtac ctgtgtggc ggatcccagc ggcgtgttgt tgtatgtcag 4080
gatgaaaatg gataaccgc aaacgactgc gtggagagaa taaaacctga tgagcaaaga 4140
gcctgtgaat ccggccctg tcctcagtgg gcttatggca actggggaga gtgcactaag 4200
ctgtgtgggt gaggcataag aacaagactg gtggctctgc agcggtccea cggtgaacgg 4260
tttccagatt tgagctgtga aattcttgat aaacctcccg atcgtgagca gtgtaacaca 4320
catgcttgtc cacacgacgc tgcattggagt actggccctt ggagctcgtg ttctgtctct 4380
tgtgtgtcag ggcataaaca acgaaatgtt tactgcatgg caaaagatgg aagccattta 4440
gaaagtgatt actgtaagca cctggctaag ccacatgggc acagaaagt ccgaggagga 4500
agatgcccc aatggaaagc tggcgcttgg agtcagtgtc ctgtgtcctg tggccgaggg 4560
gtacagcaga ggcatgtgg ctgtcagatc ggaacacaca aatagccag agagaccgag 4620
tgcaaccat acaccagacc ggagtctgaa cgcgactgcc aaggcccacg gtgtcccctc 4680
taacttggg gggcagagga atggcaagaa tgcaccaaga cctgcggcga aggtccaggg 4740
taccgcaagg tgggtgtgtg ggatgacaac aaaaacgagg tgcattggggc acgctgtgac 4800
gtgagcaagc ggccggtgga ccgtgaaagc tgtagtattg aaccctgcga gtatgtctgg 4860
atcacaggag aatggtcaga ggtaccgtcc tgggaactgt aaccatcgtc agctcagca 4920
tggcctgaga gtggcagagg gatgagtgg gggatgagtg caggaatgtg ggagacttga 4980
ggctaccgc cggatttgcc actgtgaact gtgtgtcttc tgacaagtcc tcagctttcc 5040
caagctagaa ttc

```

```

<210> SEQ ID NO:9
<211> LENGTH: 1629
<212> TYPE: PRT
<213> ORGANISM: unknown

<220> FEATURE:
<223> OTHER INFORMATION: Unknown

```

```

<400> SEQ ID NO:9
Met Gln Phe Val Ser Trp Ala Thr Leu Leu Thr Leu Leu Val Arg Asp
 1                               5                               10                               15
Leu Ala Glu Met Gly Ser Pro Asp Ala Ala Ala Ala Val Arg Lys Asp
                               20                               25                               30
Arg Leu His Pro Arg Gln Val Lys Leu Leu Glu Thr Leu Ser Glu Tyr
                               35                               40                               45
Glu Ile Val Ser Pro Ile Arg Val Asn Ala Leu Gly Glu Pro Phe Pro
                               50                               55                               60
Thr Asn Val His Phe Lys Arg Thr Arg Arg Ser Ile Asn Ser Ala Thr
 65                               70                               75                               80

```

Asp Pro Trp Pro Ala Phe Ala Ser Ser Ser Ser Ser Ser Thr Ser Thr
 85 90 95
 Gln Ala His Tyr Arg Leu Ser Ala Phe Gly Gln Gln Phe Leu Phe Asn
 100 105 110
 Leu Thr Ala Asn Ala Gly Phe Ile Ala Pro Leu Phe Thr Val Thr Leu
 115 120 125
 Leu Gly Thr Pro Gly Val Asn Gln Thr Lys Phe Tyr Ser Glu Glu Glu
 130 135 140
 Ala Glu Leu Lys His Cys Phe Tyr Lys Gly Tyr Val Asn Thr Asn Ser
 145 150 155 160
 Glu His Thr Ala Val Ile Ser Leu Cys Ser Gly Met Leu Gly Thr Phe
 165 170 175
 Arg Ser His Asp Gly Asp Tyr Phe Ile Glu Pro Leu Gln Ser Met Asp
 180 185 190
 Glu Gln Glu Asp Glu Glu Glu Gln Asn Lys Pro His Ile Tyr Arg
 195 200 205
 Arg Ser Ala Pro Gln Arg Glu Pro Ser Thr Gly Arg His Ala Cys Asp
 210 215 220
 Thr Ser Glu His Lys Asn Arg His Ser Lys Asp Lys Lys Lys Thr Arg
 225 230 235 240
 Ala Arg Lys Trp Gly Glu Arg Ile Asn Leu Ala Gly Asp Val Ala Ala
 245 250 255
 Leu Asn Ser Gly Leu Ala Thr Glu Ala Phe Ser Ala Tyr Gly Asn Lys
 260 265 270
 Thr Asp Asn Thr Arg Glu Lys Arg Thr His Arg Arg Thr Lys Arg Phe
 275 280 285
 Leu Ser Tyr Pro Arg Phe Val Glu Val Leu Val Val Ala Asp Asn Arg
 290 295 300
 Met Val Ser Tyr His Gly Glu Asn Leu Gln His Tyr Ile Leu Thr Leu
 305 310 315 320
 Met Ser Ile Val Ala Ser Ile Tyr Lys Asp Pro Ser Ile Gly Asn Leu
 325 330 335
 Ile Asn Ile Val Ile Val Asn Leu Ile Val Ile His Asn Glu Gln Asp
 340 345 350
 Gly Pro Ser Ile Ser Phe Asn Ala Gln Thr Thr Leu Lys Asn Phe Cys
 355 360 365
 Gln Trp Gln His Ser Lys Asn Ser Pro Gly Gly Ile His His Asp Thr
 370 375 380
 Ala Val Leu Leu Thr Arg Gln Asp Ile Cys Arg Ala His Asp Lys Cys
 385 390 395 400
 Asp Thr Leu Gly Leu Ala Glu Leu Gly Thr Ile Cys Asp Pro Tyr Arg
 405 410 415
 Ser Cys Ser Ile Ser Glu Asp Ser Gly Leu Ser Thr Ala Phe Thr Ile
 420 425 430
 Ala His Glu Leu Gly His Val Phe Asn Met Pro His Asp Asp Asn Asn
 435 440 445
 Lys Cys Lys Glu Glu Gly Val Lys Ser Pro Gln His Val Met Ala Pro
 450 455 460
 Thr Leu Asn Phe Tyr Thr Asn Pro Trp Met Trp Ser Lys Cys Ser Arg
 465 470 475 480
 Lys Tyr Ile Thr Glu Phe Leu Asp Thr Gly Tyr Gly Glu Cys Leu Leu
 485 490 495
 Asn Glu Pro Glu Ser Arg Pro Tyr Pro Leu Pro Val Gln Leu Pro Gly
 500 505 510
 Ile Leu Tyr Asn Val Asn Lys Gln Cys Glu Leu Ile Phe Gly Pro Gly
 515 520 525
 Ser Gln Val Cys Pro Tyr Met Met Gln Cys Arg Arg Leu Trp Cys Asn
 530 535 540
 Asn Val Asn Gly Val His Lys Gly Cys Arg Thr Gln His Thr Pro Trp
 545 550 555 560
 Ala Asp Gly Thr Glu Cys Glu Pro Gly Lys His Cys Lys Tyr Gly Phe
 565 570 575
 Cys Val Pro Lys Glu Met Asp Val Pro Val Thr Asp Gly Ser Trp Gly
 580 585 590
 Ser Trp Ser Pro Phe Gly Thr Cys Ser Arg Thr Cys Gly Gly Ile
 595 600 605

Lys Thr Ala Ile Arg Glu Cys Asn Arg Pro Glu Pro Lys Asn Gly Gly
 610 615 620
 Lys Tyr Cys Val Gly Arg Arg Met Lys Phe Lys Ser Cys Asn Thr Glu
 625 630 635 640
 Pro Cys Leu Lys Gln Lys Arg Asp Phe Arg Asp Glu Gln Cys Ala His
 645 650 655
 Phe Asp Gly Lys His Phe Asn Ile Asn Gly Leu Leu Pro Asn Val Arg
 660 665 670
 Trp Val Pro Lys Tyr Ser Gly Ile Leu Met Lys Asp Arg Cys Lys Leu
 675 680 685
 Phe Cys Arg Val Ala Gly Asn Thr Ala Tyr Tyr Gln Leu Arg Asp Arg
 690 695 700
 Val Ile Asp Gly Thr Pro Cys Gly Gln Asp Thr Asn Asp Ile Cys Val
 705 710 715 720
 Gln Gly Leu Cys Arg Gln Ala Gly Cys Asp His Val Leu Asn Ser Lys
 725 730 735
 Ala Arg Arg Asp Lys Cys Gly Val Cys Gly Gly Asp Asn Ser Ser Cys
 740 745 750
 Lys Thr Val Ala Gly Thr Phe Asn Thr Val His Tyr Gly Tyr Asn Thr
 755 760 765
 Val Val Arg Ile Pro Ala Gly Ala Thr Asn Ile Asp Val Arg Gln His
 770 775 780
 Ser Phe Ser Gly Glu Thr Asp Asp Asn Tyr Leu Ala Leu Ser Ser
 785 790 795 800
 Ser Lys Gly Glu Phe Leu Leu Asn Gly Asn Phe Val Val Thr Met Ala
 805 810 815
 Lys Arg Glu Ile Arg Ile Gly Asn Ala Val Val Glu Tyr Ser Gly Ser
 820 825 830
 Glu Thr Ala Val Glu Arg Ile Asn Ser Thr Asp Arg Ile Glu Gln Glu
 835 840 845
 Leu Leu Leu Gln Val Leu Ser Val Gly Lys Leu Tyr Asn Pro Asp Val
 850 855 860
 Arg Tyr Ser Phe Asn Ile Pro Ile Glu Asp Lys Pro Gln Gln Phe Tyr
 865 870 875 880
 Trp Asn Ser His Gly Pro Trp Gln Ala Cys Ser Lys Pro Cys Gln Gly
 885 890 895
 Glu Arg Lys Arg Lys Leu Val Cys Thr Arg Glu Ser Asp Gln Leu Thr
 900 905 910
 Val Ser Asp Gln Arg Cys Asp Arg Leu Pro Gln Pro Gly His Ile Thr
 915 920 925
 Glu Pro Cys Gly Thr Asp Cys Asp Leu Arg Trp His Val Ala Ser Arg
 930 935 940
 Ser Glu Cys Ser Ala Gln Cys Gly Leu Gly Tyr Arg Thr Leu Asp Ile
 945 950 955 960
 Tyr Cys Ala Lys Tyr Ser Arg Leu Asp Gly Lys Thr Glu Lys Val Asp
 965 970 975
 Asp Gly Phe Cys Ser Ser His Pro Lys Pro Ser Asn Arg Glu Lys Cys
 980 985 990
 Ser Gly Glu Cys Asn Thr Gly Gly Trp Arg Tyr Ser Ala Trp Thr Glu
 995 1000 1005
 Cys Ser Lys Ser Cys Asp Gly Gly Thr Gln Arg Arg Arg Ala Ile Cys
 1010 1015 1020
 Val Asn Thr Arg Asn Asp Val Leu Asp Asp Ser Lys Cys Thr His Gln
 1025 1030 1035 1040
 Glu Lys Val Thr Ile Gln Arg Cys Ser Glu Phe Pro Cys Pro Gln Trp
 1045 1050 1055
 Lys Ser Gly Asp Trp Ser Glu Cys Leu Val Thr Cys Gly Lys Gly His
 1060 1065 1070
 Lys His Arg Gln Val Trp Cys Gln Phe Gly Glu Asp Arg Leu Asn Asp
 1075 1080 1085
 Arg Met Cys Asp Pro Glu Thr Lys Pro Thr Ser Met Gln Thr Cys Gln
 1090 1095 1100
 Gln Pro Glu Cys Ala Ser Trp Gln Ala Gly Pro Trp Gly Gln Cys Ser
 1105 1110 1115 1120
 Val Thr Cys Gly Gln Gly Tyr Gln Leu Arg Ala Val Lys Cys Ile Ile
 1125 1130 1135

Gly Thr Tyr Met Ser Val Val Asp Asp Asn Asp Cys Asn Ala Ala Thr
 1140 1145 1150
 Arg Pro Thr Asp Thr Gln Asp Cys Glu Leu Pro Ser Cys His Pro Pro
 1155 1160 1165
 Pro Ala Ala Pro Glu Thr Arg Arg Ser Thr Tyr Ser Ala Pro Arg Thr
 1170 1175 1180
 Gln Trp Arg Phe Gly Ser Trp Thr Pro Cys Ser Ala Thr Cys Gly Lys
 1185 1190 1195 1200
 Gly Thr Arg Met Arg Tyr Val Ser Cys Arg Asp Glu Asn Gly Ser Val
 1205 1210 1215
 Ala Asp Glu Ser Ala Cys Ala Thr Leu Pro Arg Pro Val Ala Lys Glu
 1220 1225 1230
 Glu Cys Ser Val Thr Pro Cys Gly Gln Trp Lys Ala Leu Asp Trp Ser
 1235 1240 1245
 Ser Cys Ser Val Thr Cys Gly Gln Gly Arg Ala Thr Arg Gln Val Met
 1250 1255 1260
 Cys Val Asn Tyr Ser Asp His Val Ile Asp Arg Ser Glu Cys Asp Gln
 1265 1270 1275 1280
 Asp Tyr Ile Pro Glu Thr Asp Gln Asp Cys Ser Met Ser Pro Cys Pro
 1285 1290 1295
 Gln Arg Thr Pro Asp Ser Gly Leu Ala Gln His Pro Phe Gln Asn Glu
 1300 1305 1310
 Asp Tyr Arg Pro Arg Ser Ala Ser Pro Ser Arg Thr His Val Leu Gly
 1315 1320 1325
 Gly Asn Gln Trp Arg Thr Gly Pro Trp Gly Ala Cys Ser Ser Thr Cys
 1330 1335 1340
 Ala Gly Gly Ser Gln Arg Arg Val Val Val Cys Gln Asp Glu Asn Gly
 1345 1350 1355 1360
 Tyr Thr Ala Asn Asp Cys Val Glu Arg Ile Lys Pro Asp Glu Gln Arg
 1365 1370 1375
 Ala Cys Glu Ser Gly Pro Cys Pro Gln Trp Ala Tyr Gly Asn Trp Gly
 1380 1385 1390
 Glu Cys Thr Lys Leu Cys Gly Gly Ile Arg Thr Arg Leu Val Val
 1395 1400 1405
 Cys Gln Arg Ser Asn Gly Glu Arg Phe Pro Asp Leu Ser Cys Glu Ile
 1410 1415 1420
 Leu Asp Lys Pro Pro Asp Arg Glu Gln Cys Asn Thr His Ala Cys Pro
 1425 1430 1435 1440
 His Asp Ala Ala Trp Ser Thr Gly Pro Trp Ser Ser Cys Ser Val Ser
 1445 1450 1455
 Cys Gly Arg Gly His Lys Gln Arg Asn Val Tyr Cys Met Ala Lys Asp
 1460 1465 1470
 Gly Ser His Leu Glu Ser Asp Tyr Cys Lys His Leu Ala Lys Pro His
 1475 1480 1485
 Gly His Arg Lys Cys Arg Gly Gly Arg Cys Pro Lys Trp Lys Ala Gly
 1490 1495 1500
 Ala Trp Ser Gln Cys Ser Val Ser Cys Gly Arg Gly Val Gln Gln Arg
 1505 1510 1515 1520
 His Val Gly Cys Gln Ile Gly Thr His Lys Ile Ala Arg Glu Thr Glu
 1525 1530 1535
 Cys Asn Pro Tyr Thr Arg Pro Glu Ser Glu Arg Asp Cys Gln Gly Pro
 1540 1545 1550
 Arg Cys Pro Leu Tyr Thr Trp Arg Ala Glu Glu Trp Gln Glu Cys Thr
 1555 1560 1565
 Lys Thr Cys Gly Glu Gly Ser Arg Tyr Arg Lys Val Cys Val Asp
 1570 1575 1580
 Asp Asn Lys Asn Glu Val His Gly Ala Arg Cys Asp Val Ser Lys Arg
 1585 1590 1595 1600
 Pro Val Asp Arg Glu Ser Cys Ser Leu Gln Pro Cys Glu Tyr Val Trp
 1605 1610 1615
 Ile Thr Gly Glu Trp Ser Glu Val Pro Ser Trp Glu Leu
 1620 1625